A TEST OF THE HYPOTHESIS THAT THE SECOND PARAMETERS OF PAIRED BINOMIAL DISTRIBUTIONS ARE PAIRWISE EQUAL

by

Gene Arthur Venzke

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

THESIS

A TEST OF THE HYPOTHESIS THAT THE SECOND PARAMETERS OF PAIRED BINOMIAL DISTRIBUTIONS ARE PAIRWISE EQUAL

by

Gene Arthur Venzke

December 1968

A TEST OF THE HYPOTHESIS THAT THE SECOND PARAMETERS OF PAIRED BINOMIAL DISTRIBUTIONS ARE PAIRWISE EQUAL

by

Gene Arthur Venzke
Major, United States Army
B.S.E.E. University of Wisconsin, 1959

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 1968

ABSTRACT

A test of the hypothesis that the second parameters ("success" probabilities) of a number of paired binomial distributions are pairwise equal is derived under weak assumptions. Computer codes necessary to implement the procedure are given and a case study is used to demonstrate the procedure. Some other procedures for testing the same hypothesis under stronger assumptions are discussed and compared with the given procedure. A rapid approximate procedure is also given.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

In certain experimental situations one is interested in the hypothesis of pairwise equality of means in (say) n pairs of binomial distributions, but has no interest in what that mean (for a given pair) is. Specifically, if we denote the "success probability" parameter of the first member of the i^{th} pair by d_i and the same parameter of the second member by f_i, then it is desired to test the hypothesis

$$H_0: d_i = f_i \quad i,\ i = 1,2,\ \ldots,\ n$$

against various alternate hypotheses which are left unspecified at this point. Note that the relationships (if any) between d_i and d_j or between f_i and f_j for $i \neq j$, are of no concern, nor are the numerical values of d_i and f_i.

A practical example of the need for this type hypothesis test arises in the examination of a detection system. A basic requirement of a detection system is that it should perform better than a "random" system, which is defined as one in which d, the probability of detecting a target (given that one is actually present) is equal to f, the probability of giving a false alarm. A system may be considered better than random if $d > f$.

Suppose we have a system which may respond differently under different operators. If d_i and f_i are the detection and false alarm probabilities, respectively, for the i^{th} operator, we are naturally interested in testing

$$H_0: d_i = f_i \quad i;\ i = 1,2, \ldots, n \quad \text{against, say}$$

$$H_1: d_i > f_i \quad i;\ i = 1,2, \ldots, n.$$

In some situations, one might require a two sided test or perhaps a less stringent alternate hypothesis.

In any case, the matter of interest is whether the system performs better than "random" for each operator, without regard to possible differences between operators. For the remainder of this paper, the detection system example will serve as a "prototype" case. It is hoped that this will result in increased intuitive appeal and clarity of presentation.

A typical experiment conducted to facilitate performance of the desired hypothesis test involves n operators, the i^{th} of whom makes r_i detection attempts with targets actually present and s_i attempts with no targets present. The data then consists of n pairs of observations (j_i,k_i) where j_i is a realization of the random variable J_i, the number of true detections made by the i^{th} operator from among the r_i attempts with targets present, and k_i is a realization of the random variable K_i, the number of false alarms registered from among the s_i attempts made without

targets present. We assume the i^{th} operator makes true detections with constant probability d_i, and generates false alarms with constant probability f_i. Thus J_i is binomially distributed with "success" probability parameter d_i and number of trials parameter r_i (hereafter written $J_i \stackrel{d}{=} b(r_i,d_i)$). Similarly, we have $K_i \stackrel{d}{=} b(s_i,f_i)$.

We will assume that the experimental design and general conditions are such that the following assumptions are satisfied:

1. $r_i = s_i = m \ \forall i$ (made only for mathematical simplicity)
2. $J_i \stackrel{d}{=} b(m,d_i)$, $K_i \stackrel{d}{=} b(m,f_i) \ \forall i$; $i = 1,2, \ldots, n$
3. J_i and K_i are independent $\forall i$; $i = 1,2, \ldots, n$
4. J_i and J_j are independent for $i \neq j$
5. K_i and K_j are independent for $i \neq j$

SOME EARLIER APPROACHES

Before proceeding with our development of the desired hypothesis test, we briefly discuss two procedures which have been used in the past. Since both of these procedures require assumptions stronger than those listed above, they are of restricted applicability.

The Equal Performance Assumption. If the additional assumption that $d_i = d \ \forall i$ and $f_i = f \ \forall i$, then the experimental data may be summed into a single pair of observations $j' = \sum_{i=1}^{n} j_i$ and $k' = \sum_{i=1}^{n} k_i$ where j' and k' are, respectively, realizations of J' and K', where $J' \stackrel{d}{=} b(nm,d)$ and $K' \stackrel{d}{=} b(nm,f)$.

The natural hypotheses under test are then

$$H_0: d = f \quad \text{against} \quad H_1: d > f,$$

which may readily be tested by the exact method discussed below[1] or by an appropriate asymtotic test[2].

A hypothetical example serves to illustrate the dangers of improper application of the foregoing procedure. Suppose that n/2 of the observed pairs were (m,0), while the other n/2 were (0,m). Then $j' = nm/2$ and $k' = nm/2$ and the hypothesis that $d = f$ would be accepted. What has probably happened, however is that $d_i > f_i$ for about half the operators while $d_i < f_i$ for the remainder. Summing over all individuals has obscured information to this effect contained in the data.

In an extreme case, such as the above, the experimenter should recognize that the equal performance assumption does not hold. However suppose the case were less extreme, say that given in Table 1 (assume $m > 6$).

TABLE 1

Results of a Hypothetical Experiment

Number of observations	Value of observation (j_i, k_i)
n/4	(m - 2, 4)
n/4	(m - 6, 2)
n/4	(5, m - 3)
n/4	(1, m - 5)

Here, testing under the assumption of equal performance yields $j' = (n - 1)m/2$ and $k' = (n - 1)m/2$, and again, one would accept the hypothesis that $d = f$. With this result, the experimenter would not be likely to suspect the equal performance assumption, yet it is not clear that it is satisfied.

One might attempt to "validate" the assumption of equal performance by conducting a homogeneity[3,4] test on the observation vectors $(j_1, j_2 \ldots, j_n)$ and $(k_1, k_2 \ldots, k_n)$ or by goodness of fit tests[5] to some binomial distributions. However, the homogeneity test for binomial data is a strictly intuitive procedure, and thus not entirely desireable; also the selection of which binomial distribution to test the data against poses a considerable problem. Further, in both cases, the size of the overall procedure is difficult to determine. We therefore conclude this discussion by noting that testing for homogeneity among binomial samples is an area in which additional work appears to be needed.

The Analysis of Variance Approach. Bartlett[6] has suggested and Curtiss[7] has shown that the arcsine transformation $Y = \arcsin \sqrt{X/m}$ where $X \overset{d}{=} b(m,p)$, imparts to Y an approximate normal distribution with mean $\arcsin \sqrt{p}$ and variance $1/4m$. (Denoted $Y \overset{d}{=} N(\arcsin \sqrt{p}, 1/4)$.) Using this technique in the prototype case we have $J_i' = \arcsin \sqrt{J_i/m}$ and $K_i' = \arcsin \sqrt{K_i/m}$ distributed approximately normal with means $\arcsin \sqrt{d_i}$ and $\arcsin \sqrt{f_i}$ respectively, and the same variance $1/4m$. If we now think of having a target present as treatment D

and having no target present as treatment F, the J_i' represents a "response" under treatment D, while K_i' represents a response under treatment F. We can now model these responses, say J_i' as $J_i' = \eta_D + \phi_{Di} + \epsilon_{Di}$ where

η_D = mean (transformed) response under treatment D .

ϕ_{Di} = differential (transformed) response under treatment D for the i^{th} operator

ϵ_{Di} = (transformed) sampling error for the observation on the i^{th} operator under treatment D.

K_i' could be modeled analogously as $\eta_K + \phi_{Ki} + \epsilon_{Ki}$ and we can assume that:

1. ϕ_{ri} is asymtotically $N(0,\sigma_d^2)$ $r = K,D$; $i = 1,2,\ldots,n$
2. ϵ_{ri} is asymtotically $N(0,\sigma^2)$ $r = K,D$; $i = 1,2,\ldots,n$.

The random factor model, two way classification Analysis of Variance may then be used to test

H_0: $d_i = f_i$ $\forall i$ against the somewhat restricted and unnatural alternate hypothesis

H_1: $d_i \neq f_i$ for some i.

A further problem with this approach is that it requires m "large" and d_i and f_i not "extreme", say $0.2 < d_i < 0.8$ and $0.2 < f_i < 0.8$ in order for the arcsine transformation to yield reasonably normally distributed random variables. Since we cannot (indeed, may not want to) assure that these additional requirements are met, this approach is of limited applicability.

A MORE GENERAL APPROACH

The purpose of this paper is to develop a procedure to test the hypothesis H_0: $d_i = f_i$ $\forall i$ against various alternate hypotheses (as yet unspecified) under the four assumptions listed above, and no others.

In Chapter 2, we develop a graphical representation of the observations (j_i, k_i), derive the likelihood ratio test critical region characterization, present an exact test for H_0: $d_i = f_i$ against H_1: $d_i \neq f_i$ (or $d_i > f_i$) for a given individual, and explore three apparently unfeasible approaches.

Development of a general two stage test procedure is given in Chapter 3, and in Chapter 4 we adapt the two stage test procedure to the prototype situation, present a detailed case study, and make some comparisons with the earlier approaches. We conclude with the derivation of an approximate, but rapid procedure which may be useful in some cases. Computer codes necessary to implement the two stage procedure are given in two appendices.

CHAPTER II

BACKGROUND AND "CONVENTIONAL" APPROACHES

The user of "conventional" approaches to testing the hypothesis H_0: $d_i = f_i \ \forall i$ against some alternate meets with difficulty due to the need to specify a value of d_i (or f_i) as will be shown below. Before doing so, however, it is convenient to introduce a graphical approach to the problem, derive the likelihood ratio test, and consider a special case ($n = 1$).

A GRAPHICAL REPRESENTATION

For ease of communication, it seems advantageous to utilize a graphical display of the data such as shown in Figure 1, where the observations $(j_i k_i)$ are normalized (by dividing by m) and plotted in a unit square. Intuitively, one feels that if $d_i = f_i \ \forall i$, then the points $(j_i/m, k_i/m)$ should all lie near the line $j_i = k_i$, whereas if $d_i > f_i \ \forall i$ then the points $(j_i/m, k_i/m)$ should tend to lie to the right of the line $d_i = f_i$, etc. In a sense, the test of H_0: $d_i = f_i \ \forall i$ against some alternate hypothesis is a test of "closeness" of the points $(J_i/m, K_i/m)$ to the line $j_i = k_i$. This concept will be helpful in the succeeding discussions.

THE LIKELIHOOD RATIO TEST

It would be desireable to have a likelihood ratio test, since the Neyman-Pearson Lemma then provides that for

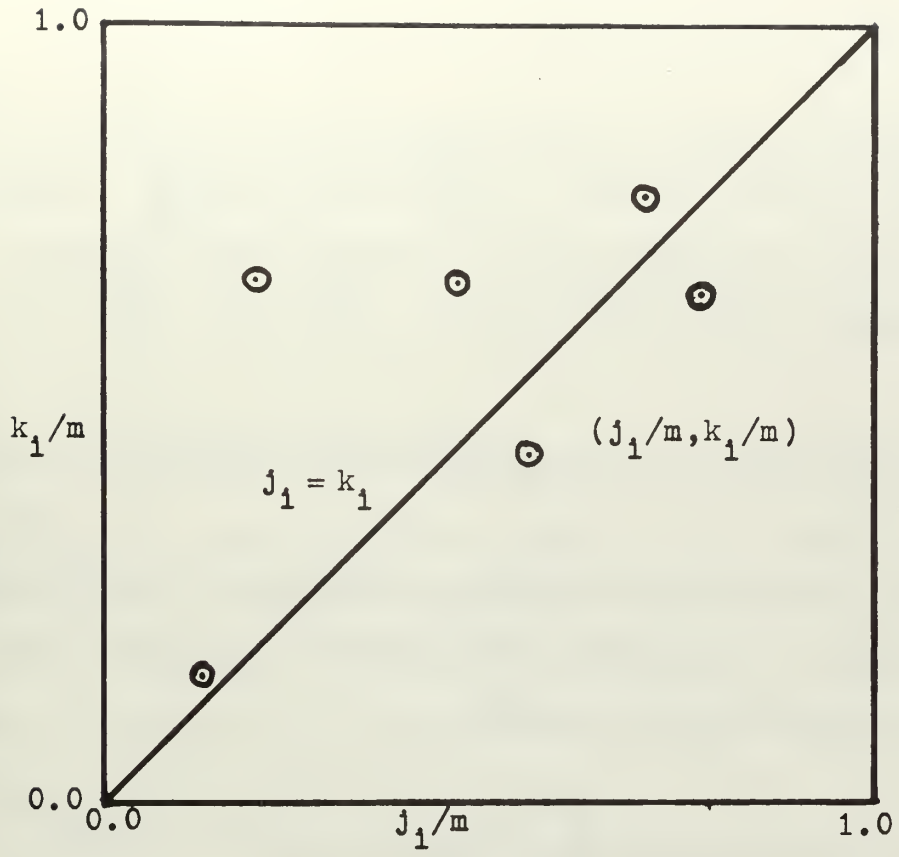

Figure 1

Graphical Representation of Data

simple hypotheses, the test is best of its size. Let us consider specifically the test of H_0: $d_i = f_i$ i against H_1: $d_i \neq f_i$ i. Let $\vec{v}$ be the vector $(j_1,k_1,j_2,k_2,\ldots,j_n,k_n)$. By independence the likelihood of v is given by

$$L(\vec{v}) = \prod_{i=1}^{n} \left\{ \binom{m}{j_i}\binom{m}{k_i}(d_i)^{j_i}(f_i)^{k_i}(1-d_i)^{m-j_i}(1-f_i)^{m-k_i} \right\} \quad (1)$$

The parameter space Ω is given by

$$\Omega = \left\{(d_i,f_i);\ 0 \leq d_i \leq 1,\ \ 0 \leq f_i \leq 1,\ i = 1,2,\ldots,n\right\}, \quad (2)$$

while the null space ω is given by

$$\omega = \left\{(d_i, f_i);\ 0 \leq d_i \leq 1,\quad 0 \leq f_i \leq 1,\quad d_i = f_i,\ i = 1,2,\ldots,n\right\}. \tag{3}$$

It is convenient to operate on Ln L rather than L directly so we maximize

$$\text{LnL} = \sum_{i=1}^{n} \left\{ j_i \ln d_i + k_i \ln f_i + (m-j_i)\ln(1-d_i) + (m-k_i)\ln(1-f_1) \right\}^* \tag{4}$$

over the null and parameter spaces. Adopting the convention $0 \cdot \text{Ln}\, 0 = 0$, (that is extending the function $x \cdot \ln x$ to 0 by continuity), and maximizing yields

$$\max_{\omega} \text{LnL}(\vec{v}) = \sum_{i=1}^{n} \left\{ (j_i + k_i)\ln\left(\frac{j_i + k_i}{2m}\right) + (2m - j_i - k_i)\ln\left(1 - \frac{j_i + k_i}{2m}\right) \right\}^* \tag{5}$$

and

$$\max_{\Omega} \text{LnL}(\vec{v}) = \sum_{i=1}^{n} \left\{ \left(j_i \ln\left(\frac{j_i}{m}\right) + k_i \ln\left(\frac{k_i}{m}\right) + (m-j_i)\ln\left(1 - \frac{j_i}{m}\right) + \right.\right.$$

$$\left.\left. (m-k_i)\ln\left(1 - \frac{k_i}{m}\right) \right\}^* . \tag{6}$$

The logarithm of the likelihood ratio $\lambda(\vec{v}) = \max_{\omega} L(\vec{v}) / \max_{\Omega} L(\vec{v})$ is given by

---

* up to additive constant terms

$$\text{Ln } \lambda(\vec{v}) = \sum_{i=1}^{n} \left\{ (j_i + k_i)\ln\left(\frac{j_i + k_i}{2m}\right) + (2m - j_i - k_i)\ln\left(1 - \frac{j_i + k_i}{2m}\right) \right.$$

$$\left. - j_i \ln\left(\frac{j_i}{m}\right) - k_i \ln\left(\frac{k_i}{m}\right) - (m - j_i)\ln\left(1 - \frac{j_i}{m}\right) - (m - k_i)\ln\left(1 - \frac{k_i}{m}\right) \right\}^{*}. \tag{7}$$

and we may characterize the test by rejecting when $\text{Ln } \lambda(\vec{v}) < C$ (constant). The boundary of the critical region is then the hypersurface in $(2n + 1)$ dimensions defined by $\text{Ln } \lambda(\vec{v}) = C$, which is difficult to visualize. However, by the independence among pairs of observations, contour lines of $\text{Ln } \lambda(\vec{v})$ in any of the n planes defined by $j_i = 0$, $k_i = 0$, $i \neq l$ for $l = 1,2,\ldots,n$, will be identical, and as shown in Figure 2. It may be noted that $\text{Ln } \lambda(\vec{v})$ could be quickly evaluated by plotting the points $(j_i/m, k_i/m)$, $i = 1,2,\ldots,n$ on Figure 2, and summing the values of the contour lines on which they fall. This reinforces our original intuitive notion of rejecting H_0 if too many points fall "far" from the diagonal $j_i = k_i$.

For a one sided test, say H_1: $d_i > f_i$ $\forall i$, the likelihood contour line plot corresponding to Figure 2 is shown on Figure 3.

It appears we are now ready to complete the procedure. We need only find C such that $P[\text{Ln } \lambda(\vec{v}) \leq C] = \alpha$ where α is the

---

* up to additive constant terms

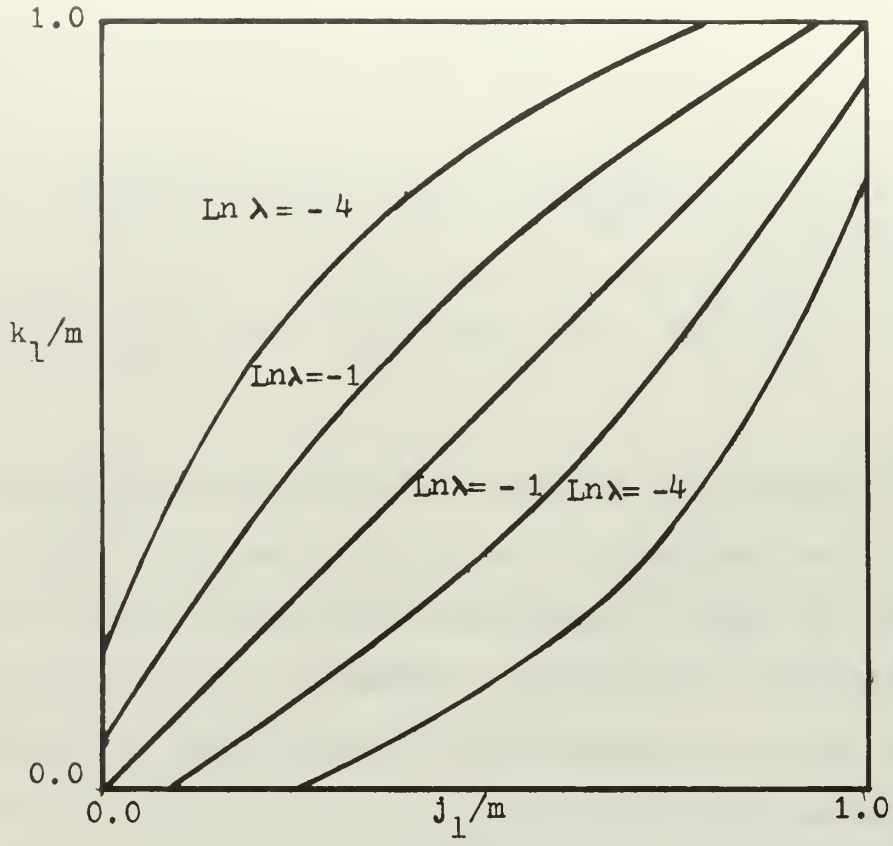

Figure 2

Contour lines of Ln (v) in the plane defined by $j_1 = 0$, $k_1 = 0$, $i \neq 1$ (two sided test)

desired size of the test. But here is where the difficulty arises. The conventional approach to finding $P[\text{Ln } \lambda(\vec{V}) \leq C]$ would be to note that

$$P[\text{Ln } \lambda(\vec{V}) = C_r] = \sum_{s \in I_r} P[\vec{V} = \vec{v}_s] \quad \text{where} \tag{8}$$

$I_r = \left\{s;\ \text{Ln } \lambda(\vec{v}_s) = C_r\right\}$, and so if $\Gamma_c = \left\{r;\ C_r \leq C\right\}$ we have

$$P[\text{Ln } \lambda(\vec{V}) \leq C] = \sum_{r \in \Gamma_c} \sum_{s \in I_r} P[\vec{V} = \vec{v}_s]. \tag{9}$$

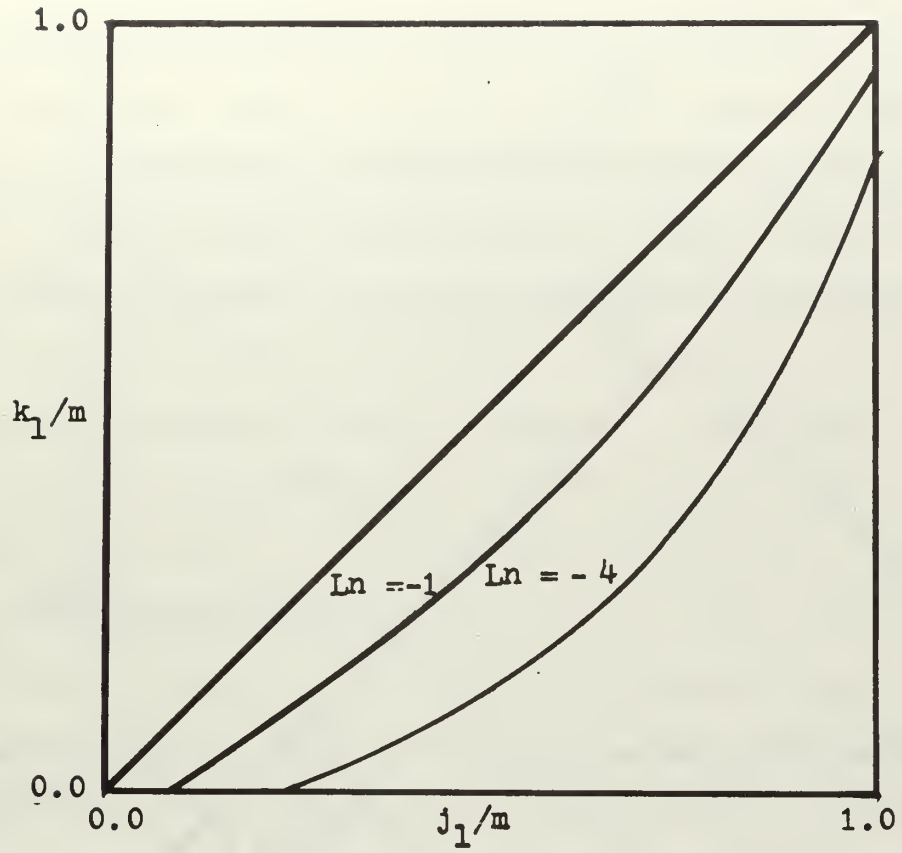

Figure 3

Contour lines of Ln (v) in the plane defined by $j_i = 0$, $k_i = 0$, $i \neq 1$ (one sided test)

Both Γ_c and I_ν are readily found for all C and ν, but the probability that $\vec{V} = \vec{v}_s$ cannot be specified, for assumption 2 states only that $J_i \stackrel{d}{=} b(m,d_i)$ and $K_i \stackrel{d}{=} b(m,f_i)$ with d_i and f_i unspecified. This inability to complete the relationship between the size of the procedure and the critical value C renders the "conventional" approach useless.

SOME "DIRECT" ATTEMPTS TO FORMULATE TESTS

Three comparatively straightforward attempts to circumvent the lack of specific distributions for J_i and K_i were

made. A brief summary of each with an explanation of why it failed follows.

The first attempt involved the transformation $u_i = (\frac{j_i}{m} - \frac{k_i}{m})\sqrt{\frac{1}{2}}$, which is a projection of the points $(j_i/m, k_i/m)$ onto an axis orthogonal to the line $j_i = k_i$, as

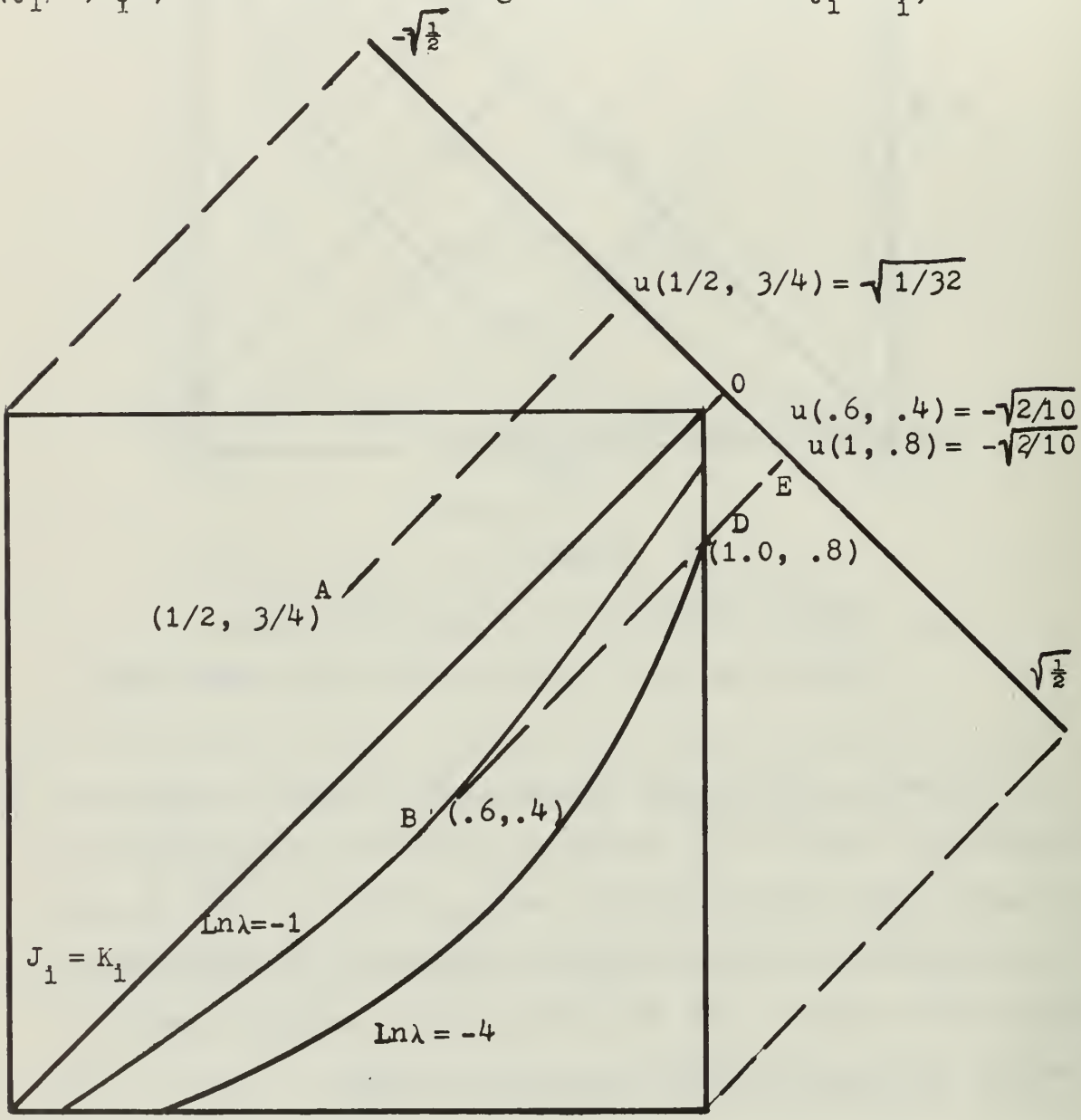

Figure 4

The Angular Transformation $u(j_i/m, k_i/m)$

illustrated in Figure 4. For example, the point A: (1/2,3/4) is transformed into $u(1/2,3/4) = \sqrt{1/32}$. Intuitively, one would hope to find a good procedure which rejects H_0 if too many of the u_i fall close to the end points of the interval $(-\sqrt{1/2}, \sqrt{1/2})$. The details of why such a procedure is not good are burdensome, but the cause is easily shown graphically by the points B and D in Figure 4. Note that both B and D map to the same point E on the U axis, yet B represents a Ln λ value of - 1, while D falls on the Ln λ = - 4 contour. Thus if C, the critical value, was - 2, say, point B should cause acceptance of H_0, while point D should cause rejection, yet given only point E on the U axis, we cannot determine whether it is the image of B or D, hence this approach fails. (More complicated transformations along the same general lines were also attempted, but without success.)

The second attempt involved use of the sums $\sum_{i=1}^{n} j_i$ and $\sum_{i=1}^{n} k_i$ as test statistics. Again, the details of the breakdown are burdensome, and a graphical argument demonstrates the problem quite clearly. As an example, we shall consider two different experiments, each with n = 2, with results plotted on Figure 5. The points A and B are from the first experiment, the points A' and B' from the second. The sums $\sum_{i=1}^{n} j_i$ and $\sum_{i=1}^{n} k_i$ are the same for each experiment (1.1 in all cases), yet clearly, the first experiment (points A and B) should tend more toward rejection than the second (points A' and B'). In short, $\sum_{i=1}^{n} j_i$ and $\sum_{i=1}^{n} k_i$ are not sufficient statistics for d_i and f_i, $i = 1,2,\ldots,n$.

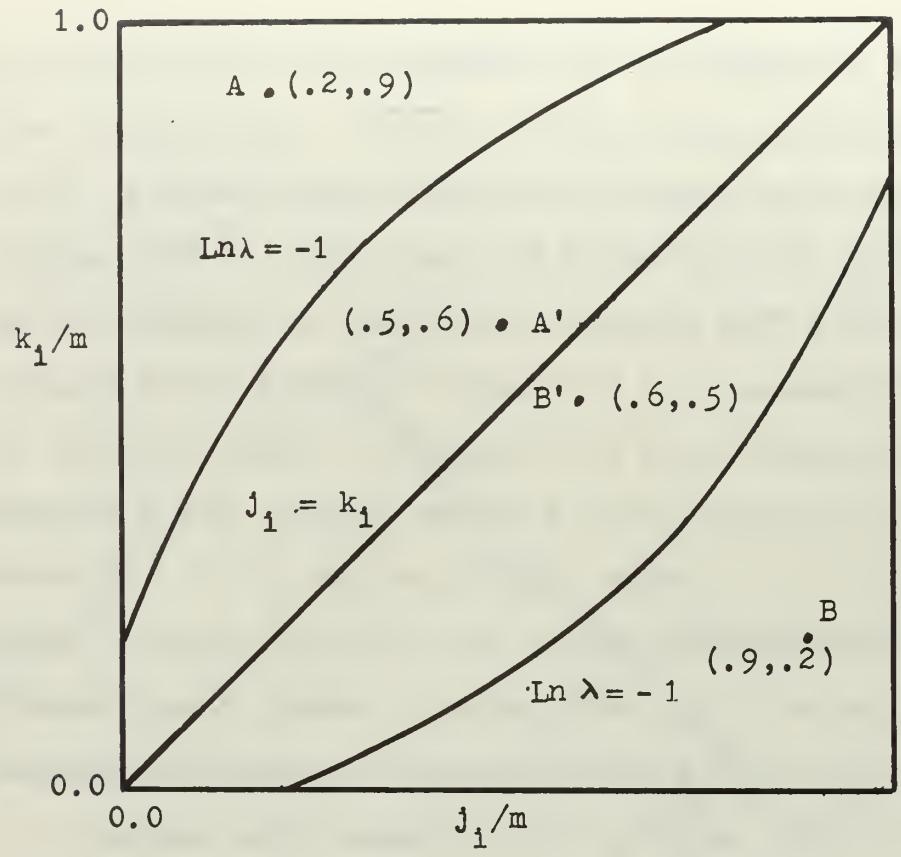

Figure 5

Two divergent experiments with equal data sums

Discussion of the third "direct" attempt is best deferred until the end of the next section, which establishes some necessary background and notation.

A SPECIAL CASE (n = 1)

When only one operator is used in the experiment, the procedure known as Fisher's Exact Test for Percentages is applicable[1]. A brief derivation follows. The test of interest is now

$$H_0: d_1 = f_1 \text{ against (say) } H_1: d_1 > f_1.$$

Now under H_0, and by independence

$$P[J_1 = j_1, K_1 = k_1] = P[J_1 = j_1] \cdot P[K_1 = k_1]$$

$$= \frac{m!m!\, d_1^{\,j_1 + k_1} (1 - d_1)^{2m - j_1 - k_1}}{j_1!(m-j_1)!k!\,(m-k_1)!} \quad . \tag{10}$$

But $P[J_1 = j_1,\ K_1 = k_1]$ is equivalent to $P[J_1 = j_1, J_1 + K_1 = j_1 + k_1]$. At this point it is convenient to introduce the random variable $L_1 = J_1 + K_1$ so that we have $P[J_1 = j_1, K_1]$ equivalent to $P[J_1 = j_1, L_1 = l_1]$ where $l_1 = j_1 + k_1$. Equation 10 may now be written

$$P[J_1 = j_1, K_1 = k_1] = \frac{m!m!d_1^{\,l_1} (1 - d_1)^{2m - l_1}}{j_1!(l_1 - j_1)!(m - j_1)!(m+j_1-l_1)!} \tag{11}$$

and we note that we must have $0 \le j_1 \le m$, $0 \le j_1 \le l_1 \le 2m$. Now (still under H_0), the random variable L_1 is the sum of two independent identically distributed binomial random variables and so is binomially distributed itself. Specifically, $L_1 \overset{d}{=} b(2m, d_1)$, so

$$P[L_1 = l_1] = \frac{(2m)!\, d_1^{\,l_1} (1 - d_1)^{2m - l_1}}{(2m - l_1)!\; l_1!} \quad . \tag{12}$$

By the definition of conditional probabilities

$$P[J_1 = j_1 \mid L_1 = l_1] = \frac{P[J_1 = j_1, L_1 = l_1]}{P[L_1 = l_1]} \quad . \tag{13}$$

Substituting (11) into (12) and (13) yields

$$P[J_1 = j_1 \mid L_1 = l_1] = \frac{\dfrac{m!m!d_1^{\,l_1}(1-d_1)^{2m-l_1}}{j_1!(l_1-j_1)!(m-j_1)!(m+j_1-l_1)!}}{\dfrac{(2m)!d_1^{\,l_1}(1-d_1)^{2m-l_1}}{(2m-l_1)!\,l_1!}}$$

$$= \frac{m!m!(2m-l_1)!\,l_1!}{(2m)!j_1!(l_1-j_1)!(m-j_1)!(m+j_1-l_1)!} \tag{14}$$

Rearranging appropriately yields

$$P[J_1 = j_1 \mid L_1 = l_1] = \frac{\binom{m}{j_1}\binom{m}{l_1-j_1}}{\binom{2m}{l_1}} \tag{15}$$

which is the hypergeometric distribution with first parameter (population size) 2m, second parameter (sample size) l_1, and third parameter (proportion) 1/2.

We now have the distribution of the number of true detections J_1 given l_1 total indicated detections, and most importantly, it is independent of d_i and f_i. Let us now develop a likelihood ratio test of H_0: $d_1 = f_1$ against some as yet unspecified alternate hypothesis. Letting $\lambda(j_1|l_1)$ denote $\max_{\omega} L(j_1|l_1) / \max_{\Omega} L(j_1|l_1)$ where $\omega = \{d_1 \in [0,1];\ d_1 = f_1\}$, $\Omega = \{d_1; 0 \le d_1 \le 1\}$, and $L(j_1|l_1)$ is the conditional likelihood of j_1 given l_1, we have

$$\lambda(j_1|l_1) = \frac{P[J_1 = j_1 \mid L_1 = l_1, H_0]}{P[J_1 = j_1 \mid L_1 = l_1, H_1]} \tag{16}$$

Now under H_1

$$P[J_1 = j_1, L_1 = l_1 \mid H_1] = \frac{m!m!d_1^{j_1} f_1^{l_1-j_1}(1-d_1)^{m-j_1}(1-f_1)^{m+j_1-l_1}}{j_1!(l_1-j_1)!(m-j_1)!(m+j_1-l_1)!} \tag{17}$$

and

$$P[L_1 = l_1 \mid H_1] = \sum_{\alpha=a}^{b} \left[\frac{m!m!d_1^{\alpha}(1-d_1)^{m-\alpha} f_1^{l_1-\alpha}(1-f_1)^{m-l_1+\alpha}}{\alpha!(m-\alpha)!(l_1-\alpha)!(m-l_1+\alpha)!} \right] \tag{18}$$

where $a = \max\left\{0 \ , \ l_1 - m\right\}$ and $b = \min\left\{m \ , l_1\right\}$.

Substituting (17) and (18) into (13) we have

$$P[J_1 = j_1 \mid L_1 = l_1, H_1] = \frac{1/\left\{j_1!(l_1-j_1)!(m-j_1)!(m-l_1+j_1)!\right\}}{\sum_{\alpha=a}^{b} \frac{\left[\frac{d_1(1-f_1)}{f_1(1-d_1)}\right]^{\alpha-j_1}}{\alpha!(m-\alpha)!(l_1-\alpha)!(m-l_1+\alpha)!}} \tag{19}$$

And substituting (14) and (19) into (16), we obtain

$$\lambda(j_1 \mid l_1) = \frac{m!m!(2m-l_1)!l_1!}{(2m)!} \sum_{\alpha=a}^{b} \left\{ \frac{\left[\frac{d_1(1-f_1)}{f_1(1-d_1)}\right]^{\alpha-j_1}}{\alpha!(m-\alpha)!(l_1-\alpha)!(m-l_1+\alpha)!} \right\} \tag{20}$$

Let us now consider a specific alternate hypothesis, say $H_1: d_1 > f_1$. We then have $\frac{d_1(1-f_1)}{f_1(1-d_1)} > 1$ and $\lambda(j_1 \mid l_1)$ is monotonically decreasing in j_1. Since we would reject H_0 if the observed test statistic $\lambda(j_1 \mid l_1)$ is sufficiently small, say less than some critical value C, we can perform an equivalent

test by rejecting H_0 if the observed value of j_1 is greater than C', where $j_1 > C'$ is equivalent to $\lambda(j_1 | l_1) < C$. We note also that this test is uniformly most powerful for H_1 since the procedure would be the same for all $d_1 > f_1$. The size of the procedure is $P[J_1 > C' | l_1, H_0]$, which can be readily computed using equation 15, and the power at a specific point (d_1, f_1) can be computed using equation 19.

For a two sided test, i.e. $H_1: d_1 \neq f_1$, a similar development shows that the optimum rejection region is characterized by the rule "reject H_0 when $j_1 < A$ or $j_1 > m - A$" where A is such that

$$P[J_1 < A | l_1, H_0] + P[J_1 > m - A | l_1, H_0] = 2P[J_1 < A | l_1, H_0] = \alpha$$

(Of course, not all values of $\alpha \in [0,1]$ are available. We assume that α is chosen from among the realizable values. Randomization to achieve arbitrary sizes may be used but we shall omit the details. To include such a consideration here would only add complication without adding substantive information.)

It now seems natural to extend the above result to cases where $n > 1$. This was the third "direct" attempt alluded to in the previous section. Specifically, it was hoped that the families of probability functions $P[\sum_{i=1}^{n} J_i \mid \sum_{i=1}^{n} l_i, H_0]$ and $P[\sum_{i=1}^{n} J_i \mid \sum_{i=1}^{n} l_i, H_1]$ could be obtained and used in a fashion analogous to the preceeding. However, whereas the function $P[J_1 | l_1, H_0]$ was independent of d_1 and

f_1, such is not the case for $P[\sum_{i=1}^{n} J_i \mid \sum_{i=1}^{n} l_i, H_0]$. The need to specify d_i and f_i, $i = 1,2,\ldots,n$, which is not permissible under our null hypothesis, forced the termination of this approach.

CHAPTER III

THE GENERAL TWO STAGE TEST

While directed at the prototype experiment, the following discussion has somewhat broader applicability than to only the specific situation discussed heretofore in this paper. For that reason, we will adopt a slightly more general notation for this chapter only. We will consider a test of the hypotheses

$$H_0: \theta_{fi} = \theta_{di} \ \forall i \quad \text{against}$$

$$H_1: \left| \theta_{fi} - \theta_{di} \right| = \Delta_i \ \forall i$$

and suppose that we have available tests T_1 thru T_n of the respective hypotheses

$$H_{0i}: \theta_{fi} = \theta_{di} \quad \text{against}$$

$$H_{1i}: \left| \theta_{fi} - \theta_{di} \right| = \Delta_i$$

for each i; $i = 1,2,\ldots,n$. We shall denote the size and power of T_i by α_i and π_i, respectively.

It seems reasonable that the outcomes of each of the tests T_1 thru T_n should contain some information about H_0 and H_1. We will develop a procedure by which the results of T_1 thru T_n (hereafter called the first stage tests) can be tested in order to make inferences about H_0 and H_1. This test on the results of T_1 thru T_n will be called the

second stage test, and the complete procedure (applying the first and second stage tests) will be called a two stage test. We begin with a special case before developing the general procedure.

IDENTICAL FIRST STAGE TESTS

Suppose that H_1 is of the form $H_1: \left| \theta_{fi} - \theta_{di} \right| = \Delta \ \forall i$. Then the respective alternate hypotheses for the first stage tests will be $H_{1i}: \left| \theta_{fi} - \theta_{di} \right| = \Delta$ for each test T_i, $i = 1,2,\ldots,n$. Now if T_1 thru T_n are identical, that is have the same operating characteristic curve, and we apply each at the same size ($\alpha_i = \alpha \ \forall i$), they each provide the same power ($\pi_i(\Delta) = \pi(\Delta) \ \forall i$). Now let us define the random variables $X_1, X_2, \ldots, X_n$ by

$$X_i = \begin{cases} 0 & \text{if } T_i \text{ results in acceptance of } H_{oi} \\ 1 & \text{if } T_i \text{ results in rejection of } H_{oi} \end{cases}$$

Under H_0 all of the H_{0i} are true, so if $Y = \sum_{i=1}^{n} X_i$ then $Y \stackrel{d}{=} b(n,\alpha)$. Similarly if H_1 is true, then all of the H_{1i} are true and here $Y \stackrel{d}{=} b(n,\pi(\Delta))$. Let us denote the (general) distribution of Y as $b(n,p)$. Then if we test

$$H_0': p = \alpha \quad \text{against} \quad H_1': p = \pi(\Delta)$$

we have actually tested H_0 against H_1 since H_0' is true if and only if the H_{0i} are true for all i and H_0 is also true if and only if all the H_{0i} are true. In similar fashion H_1' implies H_1 and conversely.

Using the well known likelihood ratio test of H_0': $p = \alpha$ against H_1': $p = \pi(\Delta)$, the procedure will be to reject H_0 if the observed value of Y exceeds a critical value y' where

$$P[Y \; y' \mid H_0'] = \sum_{k=[y']^*+1}^{n} \binom{n}{k} \alpha^k (1-\alpha)^{n-k} = \alpha' \qquad (1)$$

and the symbol $[-]^*$ denotes the greatest integer function. The test of H_0' against H_1' is, of course, the second stage test. Now since we can substitute H_0 for H_0' in the left hand expression of equation (1), we see that α' is the size of the two stage procedure. Likewise, the power of the second stage test is given by

$$P[Y > y' \mid H_1'] = \sum_{k=[y]^*+1}^{n} \binom{n}{k} \pi(\Delta)^k (1-\pi(\Delta))^{n-k} = \pi' \qquad (2)$$

and here too we can replace H_1' by H_1, so π' is the power of the two stage procedure. It is interesting that α' (the size of the two stage test) may be set independently of α, the common size of the first stage tests.

An examination of equation 2 suggests that the notation $\pi' = \pi'(\Delta, \alpha', \pi(\Delta, \alpha))$ should be used to emphasize the functional dependencies involved. We will write the equivalent form $\pi' = \pi'(\Delta, \alpha', \alpha)$. Thus, while α' may be set independently of α, the choice of both α and α' influences π'. Generally speaking, we will be given (or be willing to specify) Δ and α' so our concern would be to

choose an α that maximizes $\pi'(\Delta, \alpha', \alpha)$. This requires solution of the program

$$\text{Maximize } \pi' = \sum_{K=[y']^*+1}^{n} \binom{n}{k} [\pi(\Delta,\alpha)]^k [1-\pi(\Delta,\alpha)]^{n-k}$$

$$\text{Subject to: } \sum_{k=[y']^*+1}^{n} \binom{n}{k} \alpha^k (1-\alpha)^{n-k} = \alpha', \; 0 \leq \alpha \leq 1.$$

If the appropriate OC Curves for the second stage test are available (that is if sufficiently extensive binomial tables are available), a simple search procedure may be used to find the solution. Otherwise it appears that some iterative technique will be required. In any case we will not consider this problem further in this paper, except as it applies to the prototype case (Chapter 4).

Two further observations are appropriate. First, the two stage test may be one sided if one sided first stage tests are available. Similarly, either or both of H_0 and H_1 can be made composite provided the corresponding H_{0i} and/or H_{1i} can be made composite and with a slight modification in the definition of power. The derivation in such cases requires only minor modification of the foregoing. Second, there is no assurance that the two stage test represents a uniformly most powerful test, even when the first and second stage tests are in themselves uniformly most powerful. It appears that, in general, the two stage test may not be most powerful, although its power characteristics seem to be good. We return to this question in Chapter 4.

GENERAL FIRST STAGE TESTS

We now relax the requirement that the first stage tests $T_1, T_2, \ldots, T_n$ be identical and assume only that α_i and $\pi_i(\alpha_i)$ are specified for each T_i, $i = 1,2,\ldots,n$. We will consider testing the hypothesis

$$H_0: \ \theta_{fi} = \theta_{di} \ \forall i \quad \text{against} \quad H_1: \ |\theta_{fi} - \theta_{di}| = \Delta_i \ \forall i$$

for which the appropriate first stage tests are, for each i,

$$H_{oi}: \ \theta_{fi} = \theta_{di} \quad \text{against} \quad H_{1i}: \ |\theta_{fi} - \theta_{di}| = \Delta_i$$

If some of the first stage tests are identical, we will group them together, so we will have (say) k groups of identical tests (that is, identical within a group) to consider. (For example, if no two tests T_i, T_j; $i \neq j$ are identical, then $k = n$, whereas if all tests are identical $k = 1$.) Let $\emptyset_j$ denote the number of (identical) tests in the j^{th} group, $j = 1,2,\ldots,k$, and relabel the α_i's and π_i's so that α_j and π_j represent the common size and power, respectively, of the $\emptyset_j$ tests in group j. Now using arguments identical to those in the previous section, we have as an equivalent test of H_0 against H_1, the second stage test

$$H_0': \ \vec{p}_{\emptyset} = \vec{\alpha} \quad \text{against} \quad H_1': \ \vec{p}_{\emptyset} = \vec{\pi}.$$

Here $\vec{p}_{\emptyset}$ denotes the k component vector of success probability parameters in the joint distribution of $\vec{Y} = (Y_1, Y_2, \ldots, Y_k)$ where Y_j is the number of rejections

(of H_{0i}) from the tests in group j. (Thus $Y_j \in \{0,1,\ldots,\phi_j\}$.) The vectors $\vec{\alpha}$ and $\vec{\pi}$ denote $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ and $(\pi_1,\pi_2,\ldots,\pi_k)$, respectively. Now if α' is the size of the second stage test, we have $P\left[\text{rejection of } H_0' \mid H_0'\right] = P\left[\text{rejection of } H_0' \mid H_0\right]$ so α' is the size of the two stage procedure as well. Similarly, the power of the second stage test π', is $P\left[\text{rejection of } H_0' \mid H_0'\right]$ which is equivalent to $P\left[\text{rejection of } H_0' \mid H_1\right]$, so π' is the power of the two stage procedure also.

Let us now get the second stage procedure. By independence

$$P[Y=y] = \prod_{j=1}^{k} P[Y_j = y_j] \quad . \tag{3}$$

Under H_0', Y_j is the number of "successes" in ϕ_j trials at constant probability α_j, or $Y_j \overset{d}{=} b(\phi_j,\alpha_j)$, $j = 1,2,\ldots,k$. Similarly under H_1', we have $Y_j \overset{d}{=} b(\phi_j,\pi_j)$, $j=1,2,\ldots,k$. Then, by equation 3, under H_0'

$$L_0(y) = P[Y=y \mid H_0'] = \prod_{j=1}^{k} \left\{ \binom{\phi_j}{y_j} \alpha_j^{y_j} (1-\alpha_j)^{\phi_j - y_j} \right\} \tag{4}$$

and under H_1'

$$L_1(y) = P[Y=y \mid H_1'] = \prod_{j=1}^{k} \left\{ \binom{\phi_j}{y_j} \pi_j^{y_j} (1-\pi_j)^{\phi_j - y_j} \right\} \tag{5}$$

Thus the likelihood ratio $\lambda(\vec{y}) = L_0(\vec{y}) / L_1(\vec{y})$ is given by

$$\lambda(\vec{y}) = C \cdot \prod_{j=1}^{K} \left[\frac{\alpha_j (1 - \pi_j)}{\pi_j (1 - \alpha_j)} \right]^{y_j} \qquad (6)$$

where the constant C is $\prod_{j=1}^{k} \left[\frac{1 - \alpha_j}{1 - \pi_j} \right]^{\emptyset_j}$. In general we will have $\pi_j \geq \alpha_j \ \forall j$, hence,

$\frac{\alpha_j (1 - \pi_j)}{\pi_j (1 - \alpha_j)} \leq 1 \ \forall j$, so $\lambda(\vec{y})$ is monotonically increasing as a function of any component y_j of $\vec{y}$. Characterization of the optimum rejection region is best accomplished by considering

$$\text{Ln } \lambda(\vec{y}) = \sum_{j=1}^{k} y_j [\ln\alpha_j - \ln(1 - \alpha_j) + \ln(1 - \pi_j) - \ln \pi_j] + \ln C. \qquad (7)$$

We will reject H_0' when $\ln \lambda(\vec{y})$ is less than some critical value, say Λ. Since the quantity in the brackets $[\cdot]$ in equation 7 is negative for all j, the general tendency will be to reject H_0' when several or all of the y_j's are "large" (near $\emptyset_j$), an intuitively appealing result.

Specifically, we may label all possible outcome vectors $\vec{y}$ so that $\lambda(\vec{y}_i) \leq \lambda(\vec{y}_{i+1})$, $i = 1,2,\ldots,v-1$, where v is the number of different y's possible. ($v = \sum_{j=1}^{k} [\emptyset_j + 1]$.) Then the critical region consists of the set of vectors $\{\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_r\}$ where r is such that $\sum_{i=1}^{r} P[\vec{Y} = \vec{y}_i \mid H_0'] = \alpha'$. (We again assume α' is chosen so that it is a realizable value.) The power π' is then $\sum_{i=1}^{r} P[\vec{Y} = \vec{y}_i \mid H_1']$.

There are several schemes for reducing the work required to produce the critical region (or equivalently, the acceptance region). The most straightforward is to find $\vec{y}_1$ and compute $P[\vec{Y}=\vec{y}_1 | H_0']$. If $P[\vec{Y}=\vec{y}_1 | H_0'] < \alpha'$, find $\vec{y}_2$ and compute $P[\vec{Y}=\vec{y}_2 | H_0']$. If $\sum_{i=1}^{s} P[\vec{Y}=\vec{y}_i | H_0'] < \alpha'$, find y_3 etc., until $\sum_{i=1}^{s} P[\vec{Y}=\vec{y}_i | H_0'] = \alpha'$, at which point $s=r$ and the critical region is defined. An identical procedure may be used to define the critical region for a test of specified power by using the distribution of $\vec{Y}$ under H_1' rather than H_0'. An even faster method (in most practical cases) is to start by finding y_v and working "down" to the required size (or power) in a similar manner.

A computer code (Appendix B) has been written which can complete a problem with $n=20$, $k=5$ (5 different first stage tests) in less than two minutes on a medium sized third generation computer. This program will give the acceptance region and power of a test of some given size or if desired, will print out all or a portion of the cumulative distribution function of $\vec{Y}$ under both H_0 and H_1.

In this test, as with the identical first stage tests, there is some optimal choice of the values for α_j, $j=1,2,\ldots,k$ which will maximize π' for a given size α'. The appropriate program is

$$\text{Maximize } \pi' = \sum_{i=1}^{r} P[\vec{Y}=\vec{y}_i | \pi(\alpha)]$$

$$\text{Subject to: } \sum_{i=1}^{r} P[\vec{Y} = \vec{y}_i \mid \alpha] = \alpha' \text{ and } \lambda(\vec{y}_i) \leq \lambda(\vec{y}_{i+1})$$

$$\text{for all } i,\ i = 1,2,\ldots,v-1.$$

This is a challenging program indeed and we shall not discuss the problem here, except to state that in the prototype situation it was possible to obtain powers approaching those available under the assumption of equal performance, using an ad hoc procedure (Chapter 4).

As a final comment we note that one sided tests and tests with composite hypotheses may be made if appropriate first stage tests are available. The derivation of the specific two stage test for such cases can be completed in a manner analogous to the foregoing.

CHAPTER IV

ADAPTING THE TWO STAGE TEST TO THE PROTOTYPE SITUATION

We now consider the application of the two stage test to our prototype testing situation. We wish to test

H_0: $d_i = f_i \quad \forall i$, against (say)

H_1: $d_i - f_i = \Delta_i \quad \forall i,\ i = 1,2,\ldots,n.$

The use of the two stage procedure is facilitated by specification of the Δ_i's in such a manner that Fisher's Exact Test can be used as a first stage test. Thus we will specify Δ_i so that $\pi_i(\Delta_i)$ is a constant for all d_i and f_i which satisfy $d_i - f_i = \Delta_i$.

Let us consider equation 19 of Chapter II, which may be used to compute the power of Fisher's Exact Test. We consider now, however, the application of Fisher's test to the i^{th} individual. For the i^{th} individual, equation 19 becomes

$$P\left[J_i = j_i \mid L_i = l_i, H_1\right] =$$

$$\frac{1}{\sum_{\gamma=a}^{b} \frac{d_i(1-f_i)^{\gamma-j_i}}{f_i(1-d_i)} \cdot \frac{j_i!(l_i-j_i)!(m-j_i)!(m+j_i-l_i)!}{\gamma!(l_i-\gamma)!\ (m-\gamma)!(m+\gamma-l_i)!}} \qquad (1)$$

where $a = \max\{0,\ l_i - m\}$, and $b = \min\{m,\ l_i\}$.

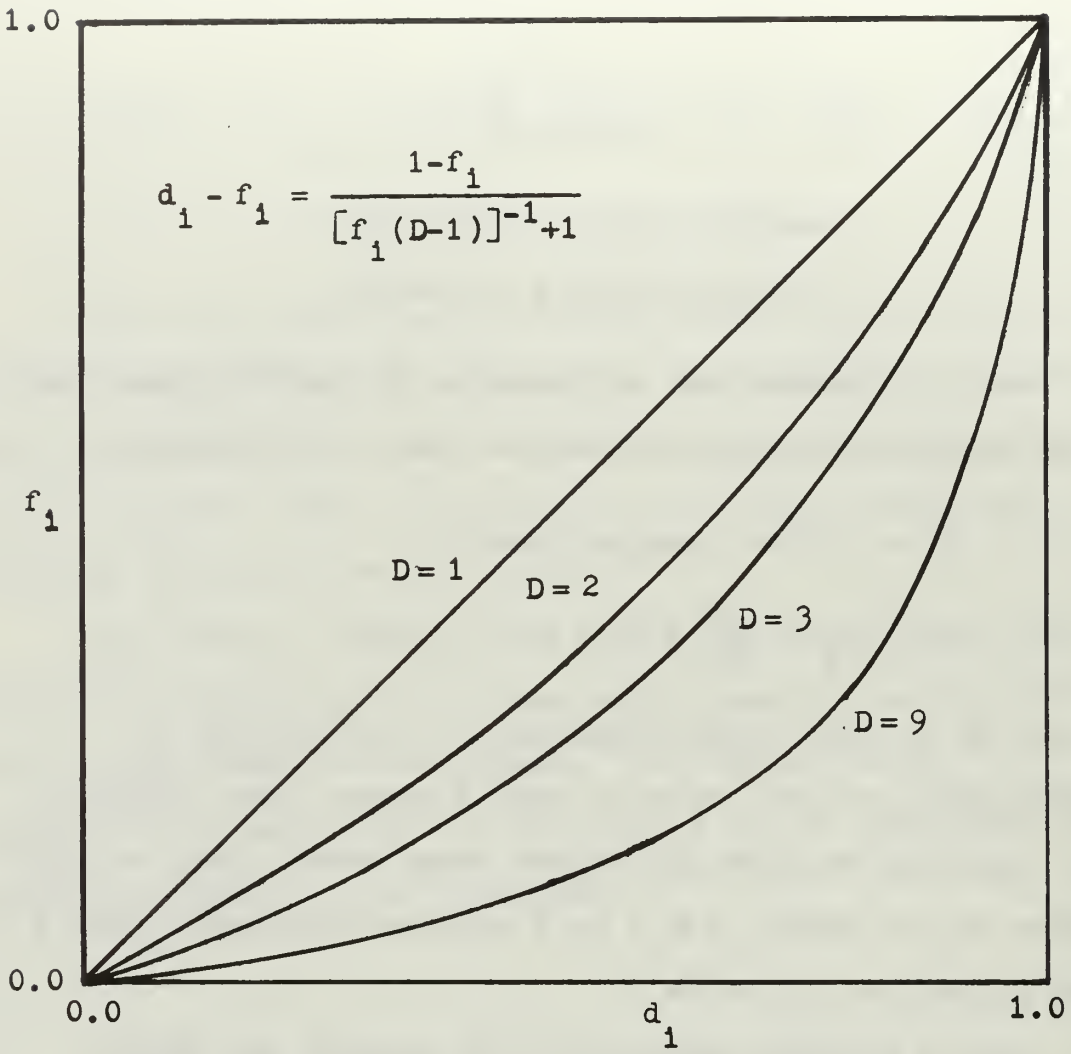

FIGURE 6

Some Sample Specifications of H_{1i}

Let us specify

$$H_{1i}: d_i - f_i = \frac{1 - f_i}{[f_i - (D-1)]^{-1}+1}$$

where $D \in (0, \infty)$.

A graphical interpretation of this specification is shown in Figure 6 for several values of D.

Two interesting features emerge from this specification. First, as seen in Figure 6, this specification for H_{1i} tends

to satisfy one's intuitive feelings about a reasonable form for H_{1i}. That is, we feel that the difference $d_i - f_i$, about which we are concerned, can be relatively large for f_i at midrange, but should shrink as f_i approaches its extreme values. Second, solving $d_i - f_i = (1-f_i)/\left[f_i(D-1)\right]^{-1} + 1$ for $\frac{d_i(1-f_i)}{f_i(1-d_i)}$ yields $\frac{d_i(1-f_i)}{f_i(1-d_i)} = D$, and equation 1 may be written

$$P[J_i = j_i \mid L_i = l_i, H_1] = \frac{1}{\sum_{r=a}^{b} \left\{ [D]^{r-j_i} \cdot \frac{j_i!(l_i-j_i)!(m-j_i)!(m+j_i-l_i)!}{r!(l_i-r)!(m-r)!(m+r-l_i)!} \right\}} \tag{2}$$

The power of the (first stage) test is then

$$\pi_i = \sum_{j_i=j_i'+a+1}^{b} \frac{1}{\sum_{r=a}^{b} D^{r-j_i} \cdot \frac{j_i!(l_i-j_i)!(m-j_i)!(m+j_i-l_i)!}{r!(l_i-r)!(m-r)!(m+r-l_i)!}} \tag{3}$$

where j_i' is such that

$$P\left[J_i > j_i \mid L_i = l_i, H_{0i}\right] = \sum_{j_i=j_i'+a+1}^{b} \frac{\binom{m}{j_i}\binom{m}{l_i-j_i}}{\binom{2m}{l_i}} = \alpha_i \tag{4}$$

Should we desire a two sided test we can specify

$$H_{1i}: \left| d_i - f_i \right| = \frac{1 - f_i}{\left[f_i(D-1)\right]^{-1} + 1}$$

Solving for $\frac{d_i(1-f_i)}{f_i(1-d_i)}$ in this specification of H_{1i} yields the solutions D and 1/D. It may be readily verified that equation 3 yields the same value of π_i when 1/D is substituted

for D. Hence, the power of this (two sided) test is still given by equation 3 provided the value of j_i' is set by

$$\alpha_i = P[J_i > j_i \mid L_1 = l_i, H_{0i}] + P[J_i < l_i - j_i \mid L_i = l_i, H_{0i}] =$$

$$2 \cdot \sum_{j_i = j_j' + a + 1}^{b} \frac{\binom{m}{j_i}\binom{m}{l_i - j_i}}{\binom{2m}{l_i}} \qquad (5)$$

THE TWO STAGE TEST FOR THE PROTOTYPE CASE

We now discuss our test of the hypotheses

$$H_0: d_i = f_i \quad \forall i \quad \text{against}$$

$$H_1: d_i - f_i = \frac{1 - f_i}{[f_i(D-2)]^{-1} + 1} \quad \forall i,\ D \in (0, \infty).$$

The first stage tests $T_1, T_2, \ldots, T_n$ will be, for each i, $i = 1,2,\ldots,n$, Fisher's Exact Test of the hypotheses

$$H_{0i}: d_i = f_i \quad \text{against} \quad H_{1i}: d_i - f_i = \frac{1 - f_i}{[f_i(D-1)]^{-1} + 1}$$

Given the observed values $\vec{l} = (l_1, l_2, \ldots, l_n)$ of $\vec{L} = (L_1, L_2, \ldots, L_n)$ (total indicated detections by each subject) we form (in general) m groups of results by the rule "group γ contains all results where $l_i = \gamma$ or $l_i = 2m - \gamma$, $\gamma = 1,2,\ldots,n$." Because Fisher's test is the same for $l_i = \gamma$ or $l_i = 2m - \gamma$, the tests to be applied within each group

are indeed identical. Continuing in the notation of Chapter 3, the number of results (or equivalently, the number of first stage tests) in group τ will be denoted ϕ_τ for $\tau = 1,2,\ldots,m$. Of course some of the ϕ_τ's may be zero, and while the derivation of the two stage test did not take into account that possibility, it is easily shown that any number of the ϕ_τ's may be zero without affecting the procedure or its power. It is convenient here to set k, the number of different tests, equal to m and allow some of the ϕ_τ's to be zero, as necessary.

We may use the computer code of Appendix A to compute the available values of α_τ and π_τ, that is, the possible non-randomized sizes and corresponding powers of the (first stage) tests in each group. We then choose, by a method discussed below, a size α_τ for the tests in each group and apply Fisher's test at that size to all results in that group, for each group $\tau, \tau = 1,2,\ldots,m$. The computer code of Appendix B may then be used to obtain a characterization of an acceptance region for the second stage test at the desired size α'. Finally we form the test statistic $\vec{Y} = (Y_1, Y_2, \ldots, Y_n)$ where Y_τ is the number of rejections of H_{0i} from among the ϕ_τ first stage tests applied to the results in group . If $\vec{Y}$ is in the acceptance region, we accept H_o, otherwise we reject H_0 and the procedure is complete.

It is interesting that the power of the two stage test just described is a random variable. This is true because

in the equation

$$\pi' = \sum_{i=1}^{r} P\left[\vec{Y} = \vec{y}_i \mid \vec{\pi}(\vec{\alpha})\right] \qquad (6)$$

the values of α_γ and $\pi_\gamma(\alpha_\gamma)$ depend upon the realization of $\vec{L}$, (or more specifically, the values of $\emptyset_\gamma$, $\gamma = 1,2,\ldots,m$). We note further that the concept of expected power cannot be applied here, since our assumptions preclude specification of a distribution for $\vec{L}$. (Indeed, suppose we did assume a distribution for $\vec{L}$. Then we could (in theory) derive the distributions of J_i and $L_i - J_i$, which would be logically equivalent to fixing values of d_i and f_i, violating our assumptions.)

However, even when the power of the two stage test obtains its "poorer" values, it still has good power for practical purposes, and in fact that power approaches the power available from tests made under the assumption of equal performance by all individuals. We next consider the choice of the best set of sizes for the first stage tests.

CHOOSING FIRST STAGE SIZES

As noted in Chapter 3, the "optimal" choice of first stage sizes α_γ, $\gamma = 1,2,\ldots,m$ would be given by a solution to the program

$$\text{Maximize} \quad \pi' = \sum_{i=1}^{r} P\left[\vec{Y} = \vec{y}_i \mid \vec{\pi}(\vec{\alpha})\right]$$

Subject to: $\sum_{i=1}^{r} P[\vec{Y} = \vec{y}_i | \vec{\alpha}] = \alpha'$ and

$$\lambda(\vec{y}_i) \leq \lambda(\vec{y}_{i+1}), \quad i = 1,2,\ldots,v-1$$

where $v = \prod_{i=1}^{m} (\phi_i + 1)$, $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and

$\vec{\pi}(\vec{\alpha}) = (\pi_1(\alpha_1), \pi_2(\alpha_2), \ldots, \pi_m(\alpha_m))$.

We have been unable to solve this program in general, so we offer only a rule of thumb for choosing the sizes of the first stage tests. This practical approach has been successful in the cases examined thus far. The rule involves use of non-randomized values for α_r only, and only assures a "near optimal" vector $\alpha_w = (\alpha_{1w}, \alpha_{2w}, \ldots, \alpha_{mw})$ where by "near optimal" we mean that only minor improvement in the value of π' (say in the second or third decimal place) would accrue if the true optimal vector $\vec{\alpha}_o$ were found.

The first step involves finding an upper bound for π'. Consider the result $\phi_r = 0, r = 1,2,\ldots,m-1$, $\phi_m = n$, which represents the most "informative" result obtainable. Since such a result reduces to the identical first stage tests situation, we would be testing the second parameter of a binomial distribution, that is

H_0: $p = \alpha_m$ against H_1: $p = \pi_m$.

The best choice of (α_m, π_m) can be obtained for that case by enumerating the powers π' obtained from each of the available (α_r, π_r) pairs, using tables of the binomial distribution. Let π_u' be the highest such power obtained for the desired size α'.

Now, using the computer code of Appendix B, we may compute π_a' for the vector α_a $(\alpha_{1a}, \alpha_{2a}, \ldots, \alpha_{ma})$ where α_{ra}, $r = 1,2,\ldots,m$ denotes the largest non-zero, non-randomized value of α_r less than or equal to 0.5. For a two stage test of size .05, if π_a' is within (say) .02 of π_u', it seems reasonable to accept α_a as "near optimal" since any other choice of α will improve π' by .02 at most. If π_a' is not within .02 of π_u', there is cause to believe it is not "near optimal" and a new vector $\vec{\alpha}_b$ should be tried. No firm suggestion for choosing $\vec{\alpha}_b$ is given, but the following observations, made from the experience of the case study to follow, are given as guidance.

Components of α for which ϕ_r is large seem to be most sensitive (that is, a change of one "step" in available α_r values causes a relatively large change in π'). Components of $\vec{\alpha}$ for which ϕ_r is small (say 1) seem quite insensitive. Thus in seeking the optimal $\vec{\alpha}$ it would seem advisable to change the α_r's for which ϕ_r is large first.

2. It appears that each component may be "near-optimized" separately. That is, if ϕ_q is the largest ϕ_r value, then one may find the best value of α_q for

arbitrary values of the other α_r's. Having found a "near optimal" value for α_q, say α_{qo}, then one can find a "near optimal" value for α_p, say α_{po}, in the vector $(\alpha_1, \alpha_2, \ldots, \alpha_p, \ldots, \alpha_{qo}, \ldots)$ where α_r is arbitrary for all r except p and q and ϕ_p is the next largest ϕ_r. This procedure may be continued until all components of $\vec{\alpha}$ have been "near-optimized". Due to interactions between components of $\vec{\alpha}$ the resulting vector may not be optimal but the case study experience provides strong evidence that it will be near optimal.

While the foregoing seems to suggest a lengthy procedure, in a very limited set of test cases considered, the initial vector $\vec{\alpha}_a$ was "near optimal" and no further work was necessary. Thus it may be that the choice of $\vec{\alpha}$ is not a problem in practice.

A CASE STUDY

Our prototype experiment was based on an experiment conducted at the US Naval Postgraduate School in the spring of 1968. There were $n = 21$ subjects, each of whom made 8 attempts to detect actual targets and 8 attempts to detect non-existent targets ($m = 8$). The value $D = 3$ was used in the specification of H_1, making the hypotheses under test

$$H_0: \; d_i = f_i \; \forall i \quad \text{against} \quad H_1: \; d_i - f_i = \frac{1 - f_i}{[2f_i]^{-1} + 1} \quad \forall i.$$

The experimental data is given in Table 2.

TABLE 2

Results of a Prototype Experiment

Subject Number (i)	Indicated Detections (l_i)	True Detections (j_i)	Subject Number (i)	Indicated Detections (l_i)	True Detections (j_i)
1	12	5	11	2	1
2	7	5	12	3	2
3	10	5	13	10	5
4	10	4	14	11	4
5	7	5	15	7	3
6	9	5	16	14	8
7	15	7	17	12	6
8	12	5	18	10	6
9	14	8	19	7	3
10	7	3	20	7	3
			21	9	4

A near optimal vector $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_8)$ was arrived at using the procedure discussed previously. It may be informative to reconstruct that procedure at this point. The (hypothetical) case $\phi_\gamma = 0$, $\gamma = 1,2,\ldots,7, \phi_8 = 21$ was used to get an upper bound π_u' on the power for size $\alpha' = .05$. The various $(\alpha_\gamma, \pi_\gamma)$ pairs available for each group γ, $\gamma = 1,2,\ldots,8$ were obtained using the computer code of Appendix A, and are reproduced in Table 3, along with a listing of the total responses(l_i's) which fell into each given group γ according to the rule given earlier (group γ contains all results $l_i = \gamma$ or $l_i = 2m - \gamma$) and the critical value for conducting the (first stage) test at the listed size α when the total response is l_i. π_u' was found to be .994 and the first trial vector α_a was (.50,.23,.50,.28, .50,.30,.50,.31). Use of $\vec{\alpha}_a$ yielded $\pi_a' = .98401$ which is near optimal. The first stage tests were then applied at the sizes in $\vec{\alpha}_a$ with results as shown in Table 4. The vector $\vec{Y} = (Y_1, Y_2, \ldots, Y_8) = (0,2,1,0,0,1,3,0)$, which may be read directly as the right hand column of Table 4, was found to be in the acceptance region and H_0 was accepted at the .05 level of significance. Note that the computer output from the trial computation for $\vec{\alpha}_a$ includes a complete description of the acceptance region, so further computation was not necessary.

ADDITIONAL RESULTS

Although not mentioned in the section on choosing a near optimal first stage size vector $\vec{\alpha}$, there is evidence

TABLE 3

Available (α_r, π_r) Pairs for the Prototype Experiment

Group Number (r)	(α_r, π_r)	l_i	Critical Value C Reject for $J_i > C$
1	(.50,.75)	1	0
		15	7
2	(.23,.53)	2	1
		14	7
3	(.50,.83)	3	1
		13	6
	(.10,.36)	3	2
		13	1
4	(.28,.68)	4	2
		12	6
	(.04,.22)	4	3
5	(.50,.86)	12	7
		5	2
		11	5
	(.14,.50)	5	3
		11	6
	(.01,.12)	5	4
		11	7
6	(.30,.74)	6	3
		10	5
	(.06,.33)	6	4
		10	6
	(.003,.05)	6	5
		10	7
7	(.50,.88)	7	3
		9	4
	(.16,.56)	7	4
		9	5
	(.02,.18)	7	5
		9	6
	(.001,.02)	7	6
		9	7
8	(.31,.75)	8	4
	(.07,.36)	8	5
	(.005,.07)	8	6
	(.0001,.003)	8	7

TABLE 4

Results of First Stage Tests

Group Number (τ)	Number of First Stage Tests in Group ($\emptyset_\tau$)	Size Used to Perform Tests (α_τ)	Number of Rejections (Y_τ)
1	1	.50	0
2	3	.23	2
3	1	.50	1
4	3	.28	0
5	1	.50	0
6	4	.30	1
7	8	.50	3
8	0	.31	0

that a rather large number of near optimal vectors ($\vec{\alpha}$) exist. In the case study just presented, an extended investigation revealed that there are at least 24 near optimal vectors for a test of size .05. (Indeed there may be more than 24, since the investigation did not enumerate all possible vectors $\vec{\alpha}$.) Table 5 presents a resumé of the near optimal vectors found. It is interesting that at least 20 first stage size vectors yield powers which differ by less than .004. Additionally, the investigation disclosed at least 26 more vectors which produce powers greater than .974 and may be considered "near-optimal".

TABLE 5

A List of Near Optimal First Stage Size Vectors

Note: $\alpha_1 = .50$, $\alpha_2 = .23$, and $\alpha_8 = .31$ in all vectors. Only the subvector $(\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7)$ is tabulated.

Subvector	π'	Subvector	π'
(.50,.28,.14,.30,.16)	.98435	(.10,.28,.01,.30,.16)	.98138
(.50,.28,.50,.30,.16)	.98419	(.10,.28,.01,.30,.50)	.98054
(.50,.28,.50,.30,.50)	.98401	(.50,.72,.14,.30,.16)	.98039
(.50,.28,.14,.30,.50)	.98391	(.50,.72,.50,.30,.16)	.98035
(.10,.28,.50,.30,.16)	.98382	(.90,.28,.86,.30,.16)	.98015
(.10,.28,.14,.30,.16)	.98360	(.50,.72,.50,.30,.50)	.97994
(.10,.28,.50,.30,.50)	.98344	(.50.,72,.14,.30,.50)	.97971
(.10,.28,.14,.30,.50)	.98288	(.90,.72,.50,.30,.16)	.97829
(.90,.28,.50,.30,.16)	.98234	(.50,.04,.14,.30,.16)	.97821
(.50,.28,.86,.30,.16)	.98232	(.50,.04,.50,.30,.16)	.97782
(.90,.28,.14,.30,.16)	.98220	(.90,.72,.14,.30,.16)	.97768
(.90,.28,.50,.30,.50)	.98205	(.50,.28,.50,.70,.16)	.97750
(.50,.28,.86,.30,.50)	.98190	(.10,.04,.50,.30,.16)	.97745
(.90,.28,.14,.30,.50)	.98174	(.50,.72,.86,.30,.50)	.97743
(.50,.28,.01,.30,.16)	.98169	(.50,.28,.50,.06,.16)	.97741
(.50,.28,.01,.30,.50)	.98161	(.50,.28,.14,.06,.16)	.97738

If we consider an experidment somewhat smaller than in the preceeding case study, it is possible to list all possible first stage size vectors (easily) and compute the resulting powers π's. Let us consider the case $n = 13$, $m = 8$, and suppose the total responses (l_i's) were such that $\emptyset_\gamma = 0, \gamma = 1,2,3,4,8$, $\emptyset_5 = 1$, $\emptyset_6 = 4$, and $\emptyset_7 = 8$. In this instance the computer code of Appendix 2 produced values of π' for all possible first stage size vectors in less than 1 minute on the IBM 360 computer. The optimal vector $\vec{\alpha}_0$ yielded a power of .949. For small experiments, or when less than 5 of the $\emptyset$'s are non zero, this complete enumeration technique is recommended. For other cases, the number of possible vectors ($\vec{\alpha}$'s) is generally rather large and computation time grows rapidly. In the case study presented, a complete enumeration would have required more than 4 hours of computer time, which is prohibitive for most users, especially in view of the small improvement in π' which may result.

COMMENTS ON POWER

In formulating the two stage test, we have weakened some of the assumptions required for the earlier techniques considered in Chapter 1. We would therefore expect some reduction in power compared to the tests which take advantage the stronger assumptions. In order to give some clarification of this point, a numerical example is now given. Suppose the experiment discussed in the case study had been

analyzed under the additional assumption that all individuals perform equally. Under that assumption, the total number of true detections $\sum_{i=1}^{21} j_i = 97$ and the total number of false alarms $\sum_{i=1}^{21} l_i - j_i = 98$ would have been considered as observations on the random variables J' and K' where $J' \overset{d}{=} b(168,d)$ and $K' \overset{d}{=} b(168,f)$. Typically, the test of H_0: $d = f$ against H_1: $d - f = \frac{1-f}{[2f]^{-1}+1}$ would be made. Using one of the usual asymtotic techniques with size .05 results in a power of .999 at the point ($d = .75$, $f = .5$) a typical point on the curve used in the alternate hypothesis for the case study. Comparing .999 with .984, the optimal power for the two stage test in the case study, we see that relaxation of the assumptions $d_i = d \ \forall i$, and $f_i = f \ \forall i$ has "cost" .015 in power. While it is for the individual experimenter to decide on the merits of this "trade off", it seems that in general, the loss of power is slight in light of the much broader applicability of the two stage test. Additionally, at the point ($d = .9$, $f = .75$) which is also on the curve $d - f = \frac{1-f}{[2f]^{-1}+1}$ the asymtotic technique yields power of .980 whereas the two stage test continues to have power of .984. At this more extreme point, the two stage test is better than the asymtotic test under the stronger assumptions.

It may also be important to set a lower bound on the power of a two stage test for a given experiment before it is run. The poorest power would emerge if all results fell in group 1, that is, if L_1 was 1 or $m - 1$ for all i. However, the result $\emptyset_1 = n$ is so unlikely that it is felt the

hypothetical result $\emptyset_2 = n$ or even $\emptyset_3 = n$ can be used to give a practical lower bound for power which is still rather conservative (perhaps even pessimistic).

In the case study, if $\emptyset_2$ was 21, the test $H_0: p = \alpha_2$ against $H_1: p = \pi_2$ would be used to determine the lower bound on π', where (α_2, π_2) is the size - power pair which gives the best power at size .05. Completing this procedure for the case example yields .91 as practical lower bound on the power of the two stage test. Of course, the actual results were much more "informative" than $\emptyset_2 = n$, and the actual power achieved was .984.

CHAPTER V

A RAPID APPROXIMATE PROCEDURE

If we strengthen assumption 1 (p. 11) by adding the statement "m is sufficiently large to assure that the arcsine transformation[6,7] provides reasonably 'normal' random variates when applied to the experimental data", (leaving all other assumptions as stated), we may then use an approximate procedure to test the hypothesis

$$H_0: d_i = f_i \ \forall i \quad \text{against} \quad H_1: d_i \neq f_i \ \forall i.$$

Only a brief derivation will be given and we make no comments about the properties of the procedure except to remark that it is easy and rapid in application as compared to the two stage test, and in all test cases thus far examined, has provided the same results (in terms of rejection of H_0) as the two stage test.

DERIVATION OF THE PROCEDURE

We shall use the terminology of Chapter 2. Under the strengthened version of assumption 1 noted above, we have (approximately)

$$\arcsin\sqrt{J_i/m} \overset{d}{=} N(\arcsin\sqrt{d_i}\ ,\ 1/4m) \ \forall i, \text{ and}$$

$$\arcsin\sqrt{K_i/m} \overset{d}{=} N(\arcsin\sqrt{f_i}\ ,\ 1/4m) \ \forall i.$$

From the well known properties of the normal distribution

$$2m \arcsin\sqrt{J_i/m} \overset{d}{=} N(\ 2m \arcsin\sqrt{d_i}\ ,\ 1/2)\ \forall\, i, \text{ and}$$
$$2m \arcsin\sqrt{K_i/m} \overset{d}{=} N(\ 2m \arcsin\sqrt{f_i},\ 1/2)\ \forall\, i.$$

Further, for each i, $i = 1,2,\ldots,n$

$$2m(\arcsin\sqrt{J_i/m} - \arcsin\sqrt{K_i/m}) \overset{d}{=} N(\ 2m\ [\arcsin\sqrt{d_i} - \arcsin\sqrt{f_i}],1).$$

Now under H_0: $f_i = d_i\ \forall\, i$, we have

$$2m\ (\arcsin\sqrt{J_i/m} - \arcsin\sqrt{K_i/m}) \overset{d}{=} N(0,1)\ \forall\, i.$$

Denoting $2m\ (\arcsin\sqrt{J_i/m}, -\ \arcsin\sqrt{K_i/m})$ by Z_i,

we have

$$(Z_i)^2 \overset{d}{=} \chi^2_1\ ;\ i = 1,2,\ldots,n, \text{ so that}$$

$$\sum_{i=1}^{n} (Z_i)^2 \overset{d}{=} \chi^2_n\ .$$

In order to test H_0: $d_i = f_i\ \forall\, i$ against H_1: $d_i \neq f_i\ \forall\, i$ at size α, we may reject H_0 if $\sum_{i=1}^{n} (Z_i)^2 \geq \chi^2_{n,\,1-\alpha}$

REFERENCES

1. Fisher, R. A. Statistical Methods for Research Workers, Thirteenth Edition, Section 21.02, Hafner, 1958.

2. Ehrenfeld, Sylvain, and Littauer, Sebastian B., Introduction to Statistical Method, pp 309 - 312, McGraw-Hill, 1964

3. Ostle, Bernard, Statistics in Research, Second Edition; pp 128, 129, The Iowa State University Press, 1963.

4. Fisher, R. A. and Mather, K., "A Linkage Test with Mice", Annals of Eugenics 7, p. 270, 1936.

5. Snedecor, G. W. Statistical Methods, Fifth Edition, The Iowa State University Press, 1956.

6. Bartlett, M. S. "The Use of Transformations" Biometrics 3 pp 39 - 52, 1947.

7. Curtiss, J. H. "On Transformations Used in The Analysis of Variance" Annals of Mathematical Statistics 14, pp 107 - 122, 1943.

APPENDIX A

A FORTRAN CODE FOR COMPUTING OPERATING CHARACTERISTIC CURVES FOR FISHER'S EXACT TEST OF PERCENTAGES

The code which follows evaluates equation 19 of Chapter 2 appropriately to form the complementary cumulative distribution functions of $J_i \mid l_i$ for each l_i, $l_i = 1,2,\ldots,2m$ under both the null and alternate hypothesis. The input parameters PD, PF, and M correspond respectively, to d_i, f_i, and m in equation 19. The values of PD and PF used as input to the code can be any point on the alternate hypothesis curve $d_i - f_i = (1 - f_i)/\left\{ [f_i(D-1)]^{-1} + 1 \right\}$. (The program is useful, of course, for any investigation of Fisher's Test whether or not related to the specific use intended here.)

A subroutine format is used to allow maximum flexibility in the application of the code. Depending on the computer used, this subroutine should perform for values of m up to at least 40. The output includes a full definition of parameters and variables and is self-explanatory.

```
      SUBROUTINE HYPER(M,PF,PD)
      P=PD*(1-PF)/PF/(1-PD)
      Q=1.0/P
      WRITE(6,4)
    4 FORMAT(1H1,4X,'DEFINITION OF SYMBOLS:',//,10X,'M IS'
     1' THE NUMBER OF DETECTION ATTEMPTS MADE UNDER EACH '
     2'TEST CONDITION.'//,10X,'L IS THE TOTAL NUMBER OF '
     3'INDICATED DETECTIONS.'//,10X,'J IS THE NUMBER OF '
     4'CORRECT DETECTIONS.'//,10X,'SO IS THE PROBABILITY '
     5'OF OUTCOMES STRICTLY GREATER THAN J UNDER HO.'//,
     610X,'S1 IS THE PROBABILITY OF OUTCOMES STRICTLY '
     7'GREATER THAN J UNDER H1.'//)
      WRITE(6,8)M,PF,PD,PF
    8 FORMAT(5X,'DISTRIBUTION FOR M='I2,'.'//,5X,'HO: '
     1'PROBABILITY OF FALSE ALARM = PROBABILITY OF TRUE '
     2'DETECTION ='F6.4,'.'//,5X,'H1: PROBABILITY OF TRUE'
     3' DETECTION ='F6.4,'.'//,9X,'PROBABILITY OF FALSE '
     4'ALARM ='F6.4,'.'//)
      WRITE(6,5)
    5 FORMAT(5X,'L'8X,'J'9X,'SO'13X,'S1'12X,'L'8X,'J'/)
      DO 50  L=1,M
      WRITE(6,7)
    7 FORMAT(1X,//)
      SO=1.0
      S1=1.0
      LL=L+1
      DO 50 JJ=1,LL
      J=JJ-1
      XJF=1.0
      IF(J-1)11,11,9
    9 DO 10 K=1,J
      XK=K
      XJF=XJF*XK
   10 CONTINUE
   11 XLJF=1.0
      LJ=L-J
      IF(LJ)14,14,12
   12 DO 13 K=1,LJ
      XK=K
      XLJF=XLJF*XK
   13 CONTINUE
   14 XMJF=1.0
      MJ=M-J
      IF(MJ)17,17,15
   15 DO 16 K=1,MJ
      XK=K
      XMJF=XMJF*XK
   16 CONTINUE
   17 XMJLF=1.0
      MJL=M+J-L
      IF(MJL-1)20,20,18
   18 DO 19 K=1,MJL
      XK=K
      XMJLF=XMJLF*XK
   19 CONTINUE
   20 W=0.0
      U=0.0
      DO 40 KKA=1,LL
      KA=KKA-1
      XAF=1.0
      IF(KA-1)23,23,21
   21 DO 22 K=1,KA
      XK=K
      XAF=XAF*XK
   22 CONTINUE
   23 MA=M-KA
      XMAF=1.0
      IF(MA-1)26,26,24
   24 DO 25 K=1,MA
      XK=K
```

```
      XMAF=XMAF*XK
   25 CONTINUE
   26 LA=L-KA
      XLAF=1.0
      IF(LA-1)29,29,27
   27 DO 28 K=1,LA
      XK=K
      XLAF=XLAF*XK
   28 CONTINUE
   29 MAL=M+KA-L
      XMALF=1.0
      PQ=1.0
      IF(MAL-1)32,32,30
   30 DO 31 K=1,MAL
      XK =K
      XMALF=XMALF*XK
   31 CONTINUE
   32 IF(KA-J)33,35,34
   33 PQ=Q**(J-KA)
      GO TO 35
   34 PQ=P**(KA-J)
   35 UA=(XJF/XAF)*(XLJF/XLAF)*(XMJF/XMAF)*(XMJLF/XMALF)
      PQ=PQ*UA
      U=U+UA
      W=W+PQ
   40 CONTINUE
      R=1.0/W
      S=1.0/U
      ML=2*M-L
      S0=S0-S
      IF(S0)60,61,61
   60 S0=0.0
   61 S1=S1-R
      IF(S1)62,63,63
   62 S1=0.0
   63 WRITE(6,6)L,J,  S0,  S1,ML,MJL
    6 FORMAT(4X,I2,7X,I2,7X,2(F10.8,5X),I2,7X,I2,//)
   50 CONTINUE
      END
```

APPENDIX B

A FORTRAN CODE FOR DETERMINING THE ACCEPTANCE REGION FOR THE SECOND STAGE TEST IN THE TWO STAGE PROCEDURE

In general, this program can be used to generate the CDF's for two joint distributions of up to 12 independent binomial distributions with first parameters given by (KPHI(1), KPHI(2),...,KPHI(12)) in both cases and second parameters given by (ALPHA(1),ALPHA(2),...,ALPHA(12)) for the first joint distribution and by (XPI(1), XPI(2),..., XPI(12)) for the second.

The program input parameters which relate to the prototype case variables are:

KPHI(I) - $\emptyset_i$ Number of first stage tests in the i^{yh} group.

ALPHA(I) - α_i Size of first stage tests in the i^{th} group.

XPI(I) - π_i Power of first stage tests in the i^{th} group.

M - k Number of groups which contain at least one test. (If desired, all groups, including those for which $\emptyset_i = 0$ can be read in to the program but this practice increases computing time. If this practice is followed, then M corresponds to m.)

Other program input parameters are:

START The largest size for which a printed value of α' and π' is desired.

SIZE The smallest size for which a printed value of α' and π' is desired.

KP The Kind of print parameter.

KP = 0 provides a complete printout of sizes and corresponding powers for all sizes between SIZE and START plus a summarized description of the acceptance region for a test of size SIZE.

KP = -1 provides only the print out of sizes and corresponding powers for all sizes between SIZE and START.

KP = 1 provides only the summarized description of the acceptance region for a test of size SIZE plus a statement of the power of that test.

NOTE: Regardless of the KP parameter specified, the output also always includes a recap of the ϕ_i's, α_i's, and π_i's for verification purposes.

In addition to reading in the above parameters, it is also necessary to insure that the dimension of KSUM, XLAM, and KX is greater than or equal to v, which can be computed by the formula $v = \prod_{i=1}^{m} (\phi_i + 1)$. With KSUM, XLAM, and KX so dimensioned, the program will operate on any experiment for which the number of subjects (n) is less than or equal to 30 and for which k, the number of non-zero ϕ_i's is less

than or equal to 12. The program is readily extended to larger problems but experience indicates that such cases cause excessive computing time and/or exceed computer storage capacity.

```
      INTEGER*2 KSUM,KX
      DIMENSION KPHI(12),A(21,21),KT(31),ALPHA(12),
     1ABAR(12),XPI(12),PIBAR(12),XL(12),XLBAR(12),LL(12),
     2ML(12),PA(21,21,12),NY(31),KFY(31),
     3XLAM(5760),KX(5760),KSUM(5760)
      READ(5,2)M,START,KP,SIZE
    2 FORMAT(I8,F8.0,I8,F8.0)
      READ(5,1)(KPHI(I),ALPHA(I),XPI(I),I=1,M)
    1 FORMAT(I8,2F8.0)
C
CC    INITIALIZE  ARRAYS AS NECESSARY
C
      DO 10 I=1,31
      NY(I)=0
      KFY(I)=0
   10 CONTINUE
      DO 11 I=1,12
      LL(I)=0
   11 CONTINUE
      DO 16 I=1,31
      KT(I)=0
   16 CONTINUE
      KYT=1
C
CC    COMPUTE AUXILIARY PARAMETERS AND ONE PLUS THE
CC    LARGEST POSSIBLE NUMBER OF SUCCESSES (KYT)
C
      DO 15 I=1,M
      ABAR(I)=1.0-ALPHA(I)
      PIBAR(I)=1.0-XPI(I)
      XL(I)=ALPHA(I)/XPI(I)
      XLBAR(I)=ABAR(I)/PIBAR(I)
      KYT=KYT+KPHI(I)
   15 CONTINUE
C
CC    SET UNNEEDED PARAMETRS TO THEIR IDLE VALUES
C
      IF(M-12)17,19,19
   17 MM=M+1
      DO 18 I=MM,12
      KPHI(I)=0
   18 CONTINUE
C
CC    COMPUTE THE NUMBER OF DIFFERENT VECTORS POSSIBLE
C
      N=1
   19 DO 3 I=1,12
      ML(I)=KPHI(I)+1
      N=N*ML(I)
    3 CONTINUE
C
CC    SET PARAMETERS NEEDED FOR ENSUING DO LOOPS
C
      M1=ML(1)
      M2=ML(2)
      M3=ML(3)
      M4=ML(4)
      M5=ML(5)
      M6=ML(6)
      M7=ML(7)
      M8=ML(8)
      M9=ML(9)
      M10=ML(10)
      M11=ML(11)
      M12=ML(12)
      WRITE(6,200)
  200 FORMAT(1H1,//)
      NPLUS=N+1
      JT=ML(1)
      DO 700 I=2,M
      IF(JT-ML(I))701,700,700
  701 JT=ML(I)
```

```
  700 CONTINUE
C
C       COMPUTE COEFFICIENTS FOR BINOMIAL PROBABILITY
C       EXPRESSIONS
C
        DO 20 I=1,JT
        XI=I
        DO 20 J=1,I
        XJ=J
        XJBAR=XI-XJ+1.0
        A(I,J)=GAMMA(XI)/(GAMMA(XJ)*GAMMA(XJBAR))
C
C       MULTIPLY COEFFICIENTS BY APPROPRIATE POWERS OF "P"
C       AND "Q". PA(I,J,K) IS THE PROBABILITY OF J SUCCESSES
C       GIVEN I TRIALS WITH SUCCESS PROBABILITY ("P") EQUAL
C       TO XPI(K).
C
        IF(J-1)5,5,6
    5   IF(I-J)7,7,8
    7   DO 21 K=1,M
        PA(I,J,K)=A(I,J)
   21   CONTINUE
        GO TO 20
    8   DO 22 KK=1,M
        IF(I-ML(KK))301,301,22
  301   PA(I,J,KK)=A(I,J)*(PIBAR(KK)**(I-J))
   22   CONTINUE
        GO TO 20
    6   IF(I-J)9,9,4
    9   DO 23 KKK=1,M
        IF(I-ML(KKK))302,302,23
  302   PA(I,J,KKK)=A(I,J)*(XPI(KKK)**(J-1))
   23   CONTINUE
        GO TO 20
    4   DO 24 KKKK=1,M
        IF(I-ML(KKKK))303,303,24
  303   PA(I,J,KKKK)=A(I,J)*(XPI(KKKK)**(J-1))*
     1(PIBAR(KKKK)**(I-J))
   24   CONTINUE
   20   CONTINUE
C
C       COMPUTE LIKELIHOOD RATIOS FOR ALL VECTORS (XLAM(L))
C
        L=0
        DO 90 J1=1,M1
        K1=J1-1
        KB1=M1-J1
        IF(K1)30,30,31
   30   B=XLBAR(1)**KB1
        GO TO 34
   31   IF(KB1)32,32,33
   32   B=XL(1)**K1
        GO TO 34
   33   B=(XLBAR(1)**KB1)*(XL(1)**K1)
   34   DO 90 J2=1,M2
        K2=J2-1
        KB2=M2-J2
        IF(K2)35,35,36
   35   IF(KB2)91,91,37
   37   C=B*XLBAR(2)**KB2
        GO TO 40
   36   IF(KB2)38,38,39
   38   C=B*XL(2)**K2
        GO TO 40
   39   C=B*(XLBAR(2)**KB2)*(XL(2)**K2)
        GO TO 40
   91   C=B
   40   DO 90 J3=1,M3
        K3=J3-1
        KB3=M3-J3
        IF(K3)41,41,42
   41   IF(KB3)92,92,43
```

```
43 D=C*XLBAR(3)**KB3
   GO TO 46
42 IF(KB3)44,44,45
44 D=C*XL(3)**K3
   GO TO 46
45 D=C*(XLBAR(3)**KB3)*(XL(3)**K3)
   GO TO 46
92 D=C
46 DO 90 J4=1,M4
   K4=J4-1
   KB4=M4-J4
   IF(K4)47,47,48
47 IF(KB4)93,93,49
49 E=D*XLBAR(4)**KB4
   GO TO 52
48 IF(KB4)50,50,51
50 E=D*XL(4)**K4
   GO TO 52
51 E=D*(XLBAR(4)**KB4)*(XL(4)**K4)
   GO TO 52
93 E=D
52 DO 90 J5=1,M5
   K5=J5-1
   KB5=M5-J5
   IF(K5)53,53,54
53 IF(KB5)94,94,55
55 BB=E*XLBAR(5)**KB5
   GO TO 58
54 IF(KB5)56,56,57
56 BB=E*XL(5)**K5
   GO TO 58
57 BB=E*(XLBAR(5)**KB5)*(XL(5)**K5)
   GO TO 58
94 BB=E
58 DO 90 J6=1,M6
   K6=J6-1
   KB6=M6-J6
   IF(K6)59,59,60
59 IF(KB6)95,95,61
61 CC=BB*XLBAR(6)**KB6
   GO TO 64
60 IF(KB6)62,62,63
62 CC=BB*XL(6)**K6
   GO TO 64
63 CC=BB*(XLBAR(6)**KB6)*(XL(6)**K6)
   GO TO 64
95 CC=BB
64 DO 90 J7=1,M7
   K7=J7-1
   KB7=M7-J7
   IF(K7)65,65,66
65 IF(KB7)96,96,67
67 DD=CC*XLBAR(7)**KB7
   GO TO 70
66 IF(KB7)68,68,69
68 DD=CC*XL(7)**K7
   GO TO 70
69 DD=CC*(XLBAR(7)**KB7)*(XL(7)**K7)
   GO TO 70
96 DD=CC
70 DO 90 J8=1,M8
   K8=J8-1
   KB8=M8-J8
   IF(K8)71,71,72
71 IF(KB8)97,97,73
73 EE=DD*XLBAR(8)**KB8
   GO TO 76
72 IF(KB8)74,74,75
74 EE=DD*XL(8)**K8
   GO TO 76
75 EE=DD*(XLBAR(8)**KB8)*(XL(8)**K8)
   GO TO 76
```

```
   97 EE=DD
   76 DO 90 J9=1,M9
      K9=J9-1
      KB9=M9-J9
      IF(K9)77,77,78
   77 IF(KB9)98,98,79
   79 FF=EE*XLBAR(9)**KB9
      GO TO 82
   78 IF(KB9)80,80,81
   80 FF=EE*XL(9)**K9
      GO TO 82
   81 FF=EE*(XLBAR(9)**KB9)*(XL(9)**K9)
      GO TO 82
   98 FF=EE
   82 DO 90 J10=1,M10
      K10=J10-1
      KB10=M10-J10
      IF(K10)83,83,84
   83 IF(KB10)99,99,85
   85 GG=FF*XLBAR(10)**KB10
      GO TO 88
   84 IF(KB10)86,86,87
   86 GG=FF*XL(10)**K10
      GO TO 88
   87 GG=FF*(XLBAR(10)**KB10)*(XL(10)**K10)
      GO TO 88
   99 GG=FF
   88 DO 90 J11=1,M11
      K11=J11-1
      KB11=M11-J11
      IF(K11)240,240,241
  240 IF(KB11)100,100,242
  242 R=GG*XLBAR(11)**KB11
      GO TO 245
  241 IF(KB11)243,243,244
  243 R=GG*XL(11)**K11
      GO TO 245
  244 R=GG*(XLBAR(11)**KB11)*(XL(11)**K11)
      GO TO 245
  100 R=GG
  245 DO 90 J12= 1,M12
      K12=J12-1
      KB12=M12-J12
      IF(K12)246,246,247
  246 IF(KB12)101,101,248
  248 RR=R*XLBAR(12)**KB12
      GO TO 251
  247 IF(KB12)249,249,250
  249 RR=R*XL(12)**K12
      GO TO 251
  250 RR=R*(XLBAR(12)**KB12)*(XL(12)**K12)
      GO TO 251
  101 RR=R
C
C     ENCODE THE VECTOR (K1,K2, ...,K12) INTO THE NUMBER L
C
  251 L=L+1
C
C     COMPUTE THE SUM OF COMPONENENTS IN THE  VECTOR
C     (K1,K2,...,K12)
C
      KY=J12+K11+K10+K9+K8+K7+K6+K5+K4+K3+K2+K1
C
C     COUNT THE NUMBER OF VECTORS WHOSE SUM IS KY
C
      NY(KY)=NY(KY)+1
C
C     ENTER THE LIKELIHOOD RATIO AND THE ENCODED VALUE
C     OF ITS ARGUMENT INTO THEIR RESPECTIVE ARRAYS
C
      XLAM(L)=RR
      KX(L)=L
```

```
   90 CONTINUE
C
C      SORT THE LIKELIHOOD RATIOS INTO AN ORDERED ARRAY
C
       NMIN=N-1
       DO 380 I=1,NMIN
       II=I+1
       DO 381 J=II,N
       IF(XLAM(I)-XLAM(J))382,381,381
  382  T=XLAM(J)
       IT=KX(J)
       XLAM(J)=XLAM(I)
       KX(J)=KX(I)
       XLAM(I)=T
       KX(I)=IT
  381  CONTINUE
  380  CONTINUE
C
C      PRINT GENERAL HEADINGS
C
       F=1.0
       Q=1.0
       WRITE(6,209)(KPHI(I),I=1,M)
  209  FORMAT(1H1,'NUMBER OF TRIALS AND ASSOCIATED SUCCESS'
      1' PROBABILITIES UNDER THE NULL AND ALTERNATE HYPOTH'
      2'ESES'///,3X,'NUMBER'/,1X,'OF TRIALS:'3X,12(I2,4X))
       IF(KP)410,410,411
  411  WRITE(6,416)
  416  FORMAT(1X,//)
       GO TO 417
  410  WRITE(6,405)
  405  FORMAT(1H+,90X,'NOTE: THE BELOW LISTED'/,89X,'CUMUL'
      1'ATIVE PROBABILITIES ARE'/,89X,'DEFINED AS THE PROB'
      2'ABILITY')
  417  WRITE(6,208)(ALPHA(I),I=1,M)
  208  FORMAT(5X,'ALPHA:'2X,12(F4.2,2X))
       IF(KP)412,412,413
  413  WRITE(6,416)
       GO TO 418
  412  WRITE(6,406)
  406  FORMAT(1H+,88X,'THAT THE VECTOR TO THE LEFT'/,89X,
      1'OR ANY BELOW IT (THAT IS,'/,89X,'WITH SMALLER LIKE'
      2'LIHOOD)')
  418  WRITE(6,207)(XPI(I),I=1,M)
  207  FORMAT(8X,'PI:'2X,12(F4.2,2X))
       IF(KP)414,414,261
  261  WRITE(6,416)
       GO TO 419
  414  WRITE(6,407)
  407  FORMAT(1H+,88X,'OBTAINS'/)
       WRITE(6,206)
  206  FORMAT(2X,'SUM OF'22X,'NUMBER OF SUCCESSES'40X,'CUM'
      1'ULATIVE'5X,'CUMULATIVE'/,1X,'SUCCESSES'79X,'PROBAB'
      2'ILITY'4X,'PROBABILITY'/,89X,'UNDER HO'7X,'UNDER H1'
      3//)
C
C      RECOVER THE VECTORS (K1,K2,...,K12) AS (LL(1),LL(2),
C      ...,LL(12)) STARTING WITH THE ONE WITH GREATEST
C      LIKELIHOOD RATIO, THUS FORMIMG THE ACCEPTANCE
C      REGION
C
  419  DO 170 L=1,N
       KU=KX(L) -1
       MB=13-M
       DO 160 I=MB,12
       K=13-I
       LL(K)=MOD(KU,ML(K))
       KU=KU/ML(K)
  160  CONTINUE
C
C      COMPUTE THE PROBABILITY MASS FOR THE  VECTOR
C      (LL(1),LL(2),...LL(12)) UNDER HO AND H1
```

```
C
      U=1.0
      DO 165 I=1,M
  420 KL=LL(I)+1
      U=U*PA(ML(I),KL,I)
  165 CONTINUE
      W=XLAM(L)*U
C
C     DECREMENT CUMULATIVE DISTRIBUTION FUNCTIONS
C
      V=F
      F=F-W
      VV=Q
      Q=Q-U
      LY=1
C
C     RECOVER SUM OF COMPONENTS OF VECTOR
C
      DO 168 I=1,12
      LY=LY+LL(I)
  168 CONTINUE
C
C     TAG THE VECTOR WHOSE ENCODED VALUE IS L WITH THE
C     VALUE OF THE SUM OF ITS COMPONENTS
C
      KSUM(L)=LY
C
C     COUNT THE NUMBER OF VECTORS SO FAR ACCOUNTED FOR
C     WHOSE SUM IS LY
C
      KFY(LY)=KFY(LY)+1
      IF(KP)262,262,263
C
C     START PRINTING THE DISTRIBUTIONS WHEN THE SIZE
C     VALUE REACHES START
C
  262 IF(START-F)263,264,264
  264 KZ=LY-1
      WRITE(6,210)KZ
  210 FORMAT(4X,I4)
      WRITE(6,408)(LL(I),I=1,M)
  408 FORMAT(1H+,14X,12(I2,4X))
      WRITE(6,409)V,VV
  409 FORMAT(1H+,88X,2(F7.5,8X),//)
C
C     STOP PRODUCING ADDITICNAL VECTORS WHEN THE DESIRED
C     SIZE VALUE IS REACHED
C
  263 IF(SIZE-F)170,170,415
  170 CONTINUE
C
C     IF CALLED FOR, PRINT SUMMARY
C
  415 IF(KP)267,171,171
  171 WRITE(6,225)V
  225 FORMAT(5X,'THE FOLLOWING IS A SUMMARIZED STATEMENT '
     1'OF THE REJECTION AND ACCEPTANCE REGIONS FOR A TEST'
     2' OF SIZE'F6.3,'.'//)
C
C     FOR THOSE VECTORS WHOSE SUM OF COMPONENTS IS I AND
C     WHICH WERE ALL IN THE ACCEPTANCE REGION ASSIGN
C     KT(I)=2 TO SO DENOTE
C
      DO 180 I=1,KYT
      NY(I)=NY(I)-KFY(I)
      IF(NY(I))181,181,182
  181 KT(I)=2
      GO TO 180
C
C     FOR THOSE VECTORS WHOSE SUM OF ELEMENTS IS I AND
C     WHICH WERE ONLY PARTIALLY ACCOUNTED FOR IN THE
C     ACCEPTANCE REGION ASSIGN KT(I)=1 TO DENOTE THAT
```

```
C      A MORE DETAILED INVESTIGATION IS REQUIRED. FOR THOSE
C      VECTORS WHOSE ELEMENTS SUM TO I AND FOR WHICH NONE
C      APPEAR IN THE ACCEPTANCE REGION SET KT(I)=0.
C
  182       IF(KFY(I))183,183,184
  183 KT(I)=0
      GO TO 180
  184 KT(I)=1
  180 CONTINUE
      WRITE(6,230)V
  230 FORMAT(6X,'ALL VECTORS WHOSE COMPONENTS SUM TO THE '
     1'FOLLOWING NUMBERS ARE IN THE ACCEPTANCE REGION FOR'
     2' A TEST OF SIZE'F6.3,'.'//)
C
C      PRINT THOSE VECTORS OF THE ACCEPTANCE REGION WHICH
C      CAN BE IDENTIFIED SIMPLY BY THE SUM OF COMPONENTS.
C
      DO 190 I=1,KYT
      K=I-1
      IF(KT(I)-1)191,191,192
  192 WRITE(6,231)K
  231 FORMAT(25X,I3,//)
      KJ=I+1
  190 CONTINUE
C
C      INVESTIGATE VECTORS FOR WHICH KT(I)=1 AND PRINT
C      THOSE WHICH ARE IN THE ACCEPTANCE REGION.
C
  191 IF(KYT+1-KJ)267,267,266
  266 DO 195 I=KJ,KYT
      IF(KT(I))197,197,194
  194 KI=I
  195 CONTINUE
  197 IF(KI-KJ)120,196,196
  196 WRITE(6,232)V
  232 FORMAT(6X,'THE ACCEPTANCE REGION FOR A TEST OF SIZE'
     1F6.3,1X,'ALSO INCLUDES THE FOLLOWING VECTORS:'//,7X,
     2'SUM OF'/,5X,'COMPONENTS'30X,'VECTOR'//)
      DO 110 I=KJ,KI
      DO 110 J=1,L
      IF(KSUM(J)-I)110,111,110
  111 KU=KX(J)-1
      DO 112 K=MB,12
      KK=13-K
      LL(KK)=MOD(KU,ML(KK))
      KU=KU/ML(KK)
  112 CONTINUE
      KZ=KSUM(J)-1
      WRITE(6,237)KZ
  237 FORMAT(10X,I2)
      WRITE(6,233) (LL(K),K=1,M)
  233 FORMAT(1H+,25X,10(I2,4X))
      WRITE(6,236)
  236 FORMAT(1X,/)
  110 CONTINUE
C
C      PRINT THE CONCLUDING STATEMENTS
C
  120 WRITE(6,234)V
  234 FORMAT(6X,'ALL OTHER VECTORS ARE IN THE REJECTION R'
     1'EGION FOR A TEST OF SIZE'F6.3,'.'//)
      WRITE(6,235)VV,V
  235 FORMAT(6X,'THIS TEST HAS POWER OF'F6.3,1X,'AT SIZE'F6.
  267 CONTINUE
      END
```

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library Naval Postgraduate School Monterey, California 93940	2
3.	Civil Schools Branch, ECP, DPD Department of the Army Washington, D.C. 20315	1
4.	Prof. Donald R. Barr Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
5.	Maj. Gene A. Venzke 7 Loch Place Del Rey Oaks, California 93940	1
6.	Library Department of Operations Analysis (Attn: Code 55) Naval Postgraduate School Monterey, California 93940	1

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

Naval Postgraduate School
Monterey, California

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

A TEST OF THE HYPOTHESIS THAT THE SECOND PARAMETERS OF PAIRED BINOMIAL DISTRIBUTIONS ARE PAIRWISE EQUAL

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Thesis

5. AUTHOR(S) *(First name, middle initial, last name)*

Gene Arthur Venzke, Major, U.S. Army

6. REPORT DATE
December 1968

7a. TOTAL NO. OF PAGES
71

7b. NO. OF REFS
7

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Naval Postgraduate School
Monterey, California 93940

13. ABSTRACT

A test of the hypothesis that the second parameters ("success" probabilities) of a number of paired binomial distributions are pairwise equal is derived under weak assumptions. Computer codes necessary to implement the procedure are given and a case study is used to demonstrate the procedure. Some other procedures for testing the same hypothesis under stronger assumptions are discussed and compared with the given procedure. A rapid approximate procedure is also given.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
HYPOTHESIS TEST						
JOINT BINOMIAL DISTRIBUTION						
TWO STAGE TEST						
FISHER'S TEST OF PERCENTAGES						
INDEPENDENT BINOMIAL DISTRIBUTIONS						